



Analyses statistiques et modèles d'inversion pour la validation des données d'humidité des sols de la mission SMOS

Delphine Leroux

► To cite this version:

Delphine Leroux. Analyses statistiques et modèles d'inversion pour la validation des données d'humidité des sols de la mission SMOS. Hydrologie. Université Paul Sabatier - Toulouse III, 2012. Français. NNT: . tel-00776536

HAL Id: tel-00776536

<https://theses.hal.science/tel-00776536>

Submitted on 15 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse III – Paul Sabatier

Discipline ou spécialité :

Hydrologie spatiale

Présentée et soutenue par

Delphine Leroux

Le 27 Septembre 2012

Titre :

Analyses statistiques et modèles d'inversion
pour la validation des données d'humidité des sols de la mission SMOS

JURY :

Pr. Daniel Braun, *professeur à l'Université Paul Sabatier de Toulouse*

Dr. Thierry Pellarin, *chargé de recherche CNRS à Grenoble*

Dr. Catherine Prigent, *directrice de recherche CNRS à l'Observatoire de Paris*

Dr. Jean-Pierre Wigneron, *directeur de recherche à l'INRA de Bordeaux*

Ecole doctorale : Sciences de l'Univers, de l'Environnement et de l'Espace (SDU2E)

Unité de recherche : Centre d'Etudes Spatiales de la Biosphère (CESBIO)

Directeur(s) de Thèse : Dr. Yann Kerr, Dr. Béatrice Berthelot et Dr. Caroline Henry

Rapporteurs :

Dr. Pierre-Louis Frison, *maître de conférences à l'Université de Paris Est – Marne la Vallée*

Dr. Ghislain Picard, *maître de conférences à l'Université Joseph Fourier de Grenoble*

Résumé

Dans un contexte de changement climatique, de nombreuses études portent sur l'observation de plusieurs paramètres physiques tels que la température du sol, la végétation ou le contenu en eau du sol. L'humidité du sol fait partie des traceurs observés pour surveiller l'évolution du cycle hydrologique sur Terre. Les satellites d'observation de la Terre contribuent à répondre à ce besoin. Parmi eux, la première mission à avoir été spécialement conçue pour l'observation de l'humidité des sols à l'échelle globale a été la mission SMOS (*Soil Moisture and Ocean Salinity*) lancée le 2 Novembre 2009.

La première partie de ma thèse a porté sur la validation du produit d'humidité SMOS à l'échelle locale puis à l'échelle globale. Pour l'échelle locale, une comparaison entre des humidités retrouvées à partir d'observations satellite (SMOS, AMSR-E, *Advanced Microwave Scanning Radiometer*, et ASCAT, *Advanced Scatterometer*), des humidités modélisées (ECMWF, *European Centre for Medium range Weather Forecast*) et des humidités mesurées sur quatre bassins versants test aux Etats-Unis. Dans un second temps, le produit d'humidité SMOS est évalué à l'échelle globale en utilisant la méthode statistique de la triple collocation pour l'année 2010. La carte globale de l'erreur SMOS est ensuite reliée à différents paramètres comme la texture du sol (sable, argile), la couverture du sol (forêt, lacs,...) ainsi que la probabilité moyenne d'occurrence d'interférence en utilisant les méthodes ANOVA (*Analysis of Variance*) et CART (*Classification and Regression Tree*).

Par la suite, le produit d'humidité SMOS a été replacé dans un contexte historique. Il existe déjà un historique de l'humidité des sols depuis 1978 à l'échelle globale dérivé à partir de nombreuses observations par satellites mais présente des discontinuités au changement de capteur notamment. Deux méthodes statistiques sont présentées afin de relier les humidités SMOS et AMSR-E sur les quatre sites test aux États-Unis : le CDF matching (*Cumulative Density Function*) et les copules. L'année 2010 a été utilisée comme référence pour la calibration de ces deux modèles et des historiques homogènes de 2003 à 2010 ont pu être ainsi créés.

Enfin la dernière partie reviendra sur l'algorithme de SMOS et sur ses innovations par rapport à ceux déjà existants. Pour cela, un algorithme plus simple (faisant plusieurs hypothèses de départ et ne prenant pas en compte l'hétérogénéité de la surface observée) est mis en place et adapté aux acquisitions multi-angulaires de SMOS. Une comparaison entre les humidités retrouvées avec ce modèle simplifié et celles de SMOS permettra de mieux comprendre les améliorations que SMOS est capable de fournir.

Abstract

In a climate change context, many studies are dealing with the observation of numerous physical parameters such as the soil temperature, the vegetation or the soil water content. The soil moisture is one of the observed variables for monitoring the evolution of the water cycle on Earth. Satellites help to meet this need. Among them, the first spatial mission specially designed for soil moisture monitoring at the global scale is the SMOS (Soil Moisture and Ocean Salinity) mission launched on November 2, 2009.

The first part of this thesis deals with the validation of the SMOS soil moisture product at the local scale and then at the global scale. A comparison between soil moistures retrieved from satellite observations (SMOS, AMSR-E, *Advanced Microwave Scanning Radiometer*, and ASCAT, *Advanced Scatterometer*), modeled soil moistures (ECMWF, *European Centre for Medium range Weather Forecast*) and measurements is realized over four watersheds in the U.S. In a second phase, the SMOS soil moisture product is evaluated at the global scale by using the triple collocation method for 2010 data. The global SMOS error map is then linked to different parameters such as the soil texture (sand, clay), the soil cover (forest, lake,...) and also the mean occurrence probability of interferences by using the methods of ANOVA (Analysis Of Variance) and CART (Classification And Regression Tree).

Thereafter, the SMOS soil moisture product is put into a temporal context. It already exists a long time series of soil moisture from 1978 at the global scale, derived from many satellite observations. Unfortunately, discontinuities occur when sensors change. Two statistical methods are presented in order to link the SMOS and the AMSR-E soil moisture datasets over the four watersheds : CDF matching (Cumulative Density Function) and the copulas. 2010 is used as the reference year for the calibration of these two statistical models and homogeneous time series are then created from 2003 to 2010.

Finally, the last part comes back to the SMOS algorithm and to his innovations compared to the already existing algorithms. A simpler algorithm (making some assumptions and not considering the heterogeneity of the observed area) has been implemented and adapted to the L-band multi-angular SMOS observations. A comparison between the retrieved soil moisture from this simple algorithm and the SMOS product has allowed us to better understand the improvements that SMOS is capable of and how to improve the SMOS product.

Remerciements

Commençons par le commencement, je tiens en tout premier lieu à remercier du fond du cœur mes parents. Merci pour votre soutien, merci d’avoir été présent tout au long de mes études et 8 ans ça peut être long ! Vous avez toujours été là pour moi et je ne vous en remercierai jamais assez ! Comme je l’ai dit lors de mes remerciements après ma soutenance, ma maman m’avait rappelé que je voulais être chercheur sur Mars quand j’étais toute petite...encore un peu de boulot et mon rêve devrait devenir réalité... Mes premiers remerciements vont également à ma moitié, Alex. Merci d’avoir été là au quotidien, tu as été celui qui me re-motivait ou me réconfortait dans les moments difficiles et également le premier à savoir quand ça allait bien ! Ton soutien indéfectible a été ma force tout au long de ces trois années. C’est quelque part notre thèse à tous les deux !

Je tiens à remercier toutes les personnes du labo au CESBIO. Vous m’avez très bien accueillie et très vite intégrée dès mon début de stage en Janvier 2009. Je me suis toujours sentie très à l’aise au CESBIO et c’est grâce à vous et à votre gentillesse. Je remercie également tous les membres de VEGA Technologies, maintenant membre du groupe Telespazio France, pour leur accueil et leur disponibilité et plus particulièrement à l’équipe Algorithme Scientifique (Caroline, Gautier, Jérôme H., Jérôme L.) et Béatrice Berthelot maintenant à Magellium. Merci pour vos conseils, surtout en fin de thèse lors des répétitions avec Cyril Gindre et Bruno, qui m’ont été très utiles pour la soutenance finale.

Passons maintenant à l’équipe SMOS, la secte SMOS ou plutôt ma famille SMOS ! Merci à cette équipe formidable dans laquelle je me suis épanouie pendant ces trois ans, aussi bien professionnellement que personnellement. J’ai tellement appris à vos côtés, que ce soit lors de réunions ou lors de repas partagés ensemble à la cantine le midi ou au restaurant le soir, sans oublier les nombreuses conférences et autres sorties. Un grand merci à Yann K., Arnaud M., Philippe R., François C. (aka David Gueta’s BFF), Simone B., Heather L., Ahmad A., Elsa J., Lucie B., Yann S., Olivier M., Ali K., tous ces gens géniaux que j’ai eu la chance côtoyer tous les jours pendant ma thèse ! Sans oublier les cousins de la famille SMOS à Bordeaux, Jean-Pierre W., Nathalie N., merci à vous de venir nous voir de temps en temps au CESBIO avec toujours de très bonnes soirées après les réunions !

Parce qu’au boulot, il n’y a pas que le travail, certains collègues sont devenus des amis. En premier lieu, Arnaud, je n’oublierai pas tous ces moments passés avec toi et notamment notre magnifique escale à Dubaï, je crois que je m’en souviendrai toute ma vie : 4 heures dans le vieux Dubaï...charmant n’est pas l’adjectif qui conviendrait le mieux...mais plutôt complètement flippant !!! Sans oublier ces quelques semaines incroyables passées en Australie à chercher de

l'eau dans le désert et chasser les kangourous et autres émeus. Merci pour tes conseils, ton soutien et ton amitié. Émilie, si Pierre n'avait jamais été dans mon bureau, on ne se connaîtrait peut être pas...le hasard fait bien les choses ! J'ai adoré les moments passés avec toi à discuter de tout sauf du boulot, enfin peut être qu'à de rares exceptions on parlait bien de 2-3 bricoles sur Matlab quand même ! Et cette conférence EGU 2012 à Vienne qu'on s'était promis de partager ensemble, on l'a finalement fait ! Ah ces petits chocolats Mozart présents à chaque coin de rue...mais les biscuits Manner c'est quand même ce qu'il y a de meilleur !! Mais n'oublions pas ce vieil adage qui dit : **c'est formateur !** Merci d'avoir eu l'idée de la carte géante pour ma soirée de fin de thèse !!! C'est tellement une giga idée !!! Je la regarde encore quasiment tous les jours et je trouve encore des ptits mots que je n'avais pas vu la première fois, vous avez vraiment écrit partout !! Bon courage pour ta fin de thèse, tu vas tout déchirer !!

Un grand merci également à mon co-bureau Jonas avec qui j'ai donc partagé ces trois ans et même si nous n'avions pas le même rythme de travail, la permanence au bureau 104 a presque toujours été assurée ! D'ailleurs j'ai toujours le carton géant qu'on devait transformer en épouvantail géant...oui, les idées de début de thèse ne sont pas toujours les meilleures... Un grand merci à Rémi et à Christine, deux autres thésards qui m'ont accompagnés et à qui je souhaite bonne chance et bon courage pour vos soutenances qui arrivent à grand pas !!! Merci aussi à Stéphane, Amanda, Morgan, Yann S., Malika, Sandrinou, j'ai passé de très bons moments/soirées à vos côtés.

Pendant ma thèse, j'ai eu la chance de rencontrer de nombreuses personnes d'horizons très diverses. J'ai eu la chance de participer à plusieurs campagnes de mesures internationales. Au delà du travail qui était demandé, j'ai pu ainsi rencontrer de formidables personnes qui sont maintenant de vraies amies. Je pense notamment à Simone, qui fait désormais partie de l'équipe SMOS au CESBIO, et également à SändyChen et Jet avec qui j'ai partagé tellement de choses en Australie mais aussi au Danemark, en Autriche, en France, partout finalement. La recette du girl's breakfast restera dans ma mémoire : Mudshake, banane, nutella et du pain ; simple, efficient and fun ! On passe toujours de bons moments ensemble et je vous aime fort !!!!

3 ans c'est la durée de ma thèse mais c'est aussi l'âge de mon nouveau genou ! J'ai commencé ma thèse avec un genou peu fiable et maintenant que je suis de nouveau opérationnelle après 3 années de travail sur ce genou, je souhaite remercier ma team kiné (Claire, Sarah, Seb et Max) et ma team sport (Elsa, Sylvia). Cela peut paraître étrange mais j'ai passé tellement de temps à la rééducation de ce genou que cela fait intégralement partie de ce processus de réussite car quand on se sent mal dans son corps, on se sent mal dans sa tête ! Donc merci à vous de m'avoir aidée et de m'avoir rendu la confiance que j'avais égarée en chemin.

Le mot de la fin, le plus important peut être, revient à mon directeur de thèse, Yann Kerr. Merci mille fois pour mille raisons différentes. Même si au début de ma thèse, qui correspond également au lancement de SMOS, il était compliqué de se voir, tu as toujours su être disponible pour moi. Tes conseils tout au long de ma thèse m'ont toujours fait progresser et avancer dans le bon sens. Tu m'a fait confiance et tu as su m'aiguiller tout au long de ces trois ans. Tu as été un véritable papa scientifique qui a commencé et devrait poursuivre mon éducation en terme de whisky écossais car il y a encore du boulot (même si des progrès sont à noter...). Vraiment, merci pour tout.

Sommaire

1	Introduction générale	1
1.1	Contexte	2
1.1.1	Le changement climatique	2
1.1.2	L'eau sur Terre	2
1.1.2.1	Quelle quantité d'eau y a-t-il sur Terre?	2
1.1.2.2	Qu'est ce qu'une quantité d'eau dans le sol?	2
1.1.3	Importance de l'humidité du sol dans le cycle global de l'eau	4
1.1.3.1	Le cycle de l'eau	4
1.1.3.2	Le rôle de l'humidité du sol dans le cycle de l'eau	4
1.1.3.3	Mesurer l'humidité du sol	5
1.1.4	Les satellites à l'échelle globale	6
1.2	Objectifs de la thèse	8
1.3	Structure de la thèse	8
2	La télédétection et la mission SMOS	11
2.1	Théorie du signal micro-onde	12
2.1.1	Qu'est ce qu'une onde électromagnétique?	12
2.1.2	Comment cela se mesure-t-il?	13
2.1.2.1	Les imageurs micro-ondes passifs	14
2.1.2.2	Les capteurs micro-ondes actifs	14
2.1.3	Comment retrouver l'humidité du sol à partir d'un signal?	15
2.1.3.1	Notion de température de brillance pour les micro-ondes passives	15
2.1.3.2	Retrouver l'humidité à partir des températures de brillance	16
2.2	La mission SMOS	17
2.2.1	L'instrument	17
2.2.2	Les données SMOS	19
2.2.3	L'algorithme d'inversion des données SMOS	20
3	Évaluation du produit SMOS à l'échelle locale	25
3.1	Introduction	26
3.1.1	Objectifs et défis	26
3.1.2	Données satellites	27
3.1.2.1	L'instrument AMSR-E	27
3.1.2.2	L'instrument ASCAT	27
3.1.3	Données du modèle ECMWF	28
3.2	Présentation des quatre bassins versants tests aux Etats-Unis	28

3.2.1	Walnut Gulch	30
3.2.2	Little Washita	31
3.2.3	Little River	31
3.2.4	Reynolds Creek	31
3.3	Évaluation des données SMOS avec les données terrain et d'autres produits d'humidité	32
3.3.1	Présentation des données	32
3.3.2	Pré-traitement des données	32
3.3.3	Méthodologie	32
3.3.4	Comparaison avec redimensionnement linéaire du produit ASCAT	33
3.3.4.1	Walnut Gulch	33
3.3.4.2	Little Washita	35
3.3.4.3	Little River	38
3.3.4.4	Reynolds Creek	40
3.3.4.5	Différences entre les résultats du matin et de l'après-midi	42
3.3.5	Comparaison par redimensionnement statistique du produit ASCAT . . .	43
3.3.6	Dépendance d'AMSR-E à la saisonnalité	44
3.3.7	Sensibilité de VUA	46
3.4	Aperçu des produits de végétation SMOS sur les bassins versants tests	48
3.5	Conclusions et perspectives	52
4	Validation des données SMOS à l'échelle globale et sources possibles des erreurs SMOS	55
4.1	Introduction	56
4.1.1	Quelle méthode pour une étude globale ?	56
4.1.2	Objectifs	56
4.1.3	Jeux de données utilisés	57
4.2	Distribution spatiale de l'erreur variable SMOS à l'échelle globale	58
4.2.1	Méthode de triple collocation	58
4.2.1.1	Principe	58
4.2.1.2	Théorie	59
4.2.2	Conditions d'application et méthodologie	62
4.2.3	Résultats	64
4.2.3.1	SMOS / AMSR-E(VUA) / ECMWF	64
4.2.3.2	SMOS / AMSR-E(NSIDC) / ASCAT	67
4.2.3.3	SMOS / AMSR-E(VUA) / ASCAT	70
4.2.3.4	Distribution spatiale de l'erreur variable de SMOS	70
4.3	Analyse régionale et globale de l'erreur variable de SMOS	73
4.3.1	Présentation des paramètres physiques et algorithmiques possiblement liés à l'erreur variable de SMOS	73
4.3.2	Analyse de la variance (ANOVA)	75
4.3.2.1	Principe et théorie	75
4.3.2.2	Résultats	75
4.3.3	Classification des erreurs	78
4.3.3.1	Principe et théorie	78
4.3.3.2	Résultats	79
4.4	Conclusions et perspectives	81

5	Construction d'un historique de l'humidité du sol à partir de SMOS	85
5.1	Introduction	86
5.1.1	Etat de l'art	86
5.1.2	Objectif	88
5.1.3	Différences de temps et de fréquence d'acquisition	89
5.2	Méthodes statistiques pour créer une série temporelle homogène	90
5.2.1	CDF matching	91
5.2.2	Méthode des copules	91
5.2.2.1	Théorie générale	92
5.2.2.2	Familles de copules	92
5.2.2.3	Comment choisir la famille la plus adéquate aux données	93
5.2.2.4	Approche utilisée pour simuler des données	94
5.2.2.5	Exemple simple d'application	96
5.2.3	Methodologie générale	97
5.3	Séries temporelles homogènes d'humidité sur les quatre bassins versants tests aux USA	97
5.3.1	Application sur l'année 2010 des deux méthodes	97
5.3.2	Application aux données 2003-2010	99
5.4	Développements pour étendre ces méthodes à l'échelle globale	102
5.5	Conclusions	105
6	Création d'un algorithme simplifié SMOS	109
6.1	Introduction	110
6.1.1	Contexte	110
6.1.2	Objectifs	110
6.2	Modèle de transfert radiatif	111
6.2.1	Modèle de transfert radiatif à base de couches	111
6.2.1.1	Modèle de transfert radiatif à trois couches	111
6.2.1.2	Constante diélectrique du sol	113
6.2.1.3	Réflexivité du sol	114
6.2.1.4	Rugosité	114
6.2.1.5	Effets de la végétation	115
6.2.1.6	Température effective	117
6.2.2	Le modèle LPRM et les adaptations pour les acquisitions SMOS	119
6.3	Comparaison du modèle simplifié avec le modèle complexe SMOS	122
6.3.1	Paramétrisation du modèle (modèles A, B et C)	122
6.3.1.1	Profil angulaire de températures de brillance	122
6.3.1.2	Variation spatiale	125
6.3.1.3	Variation temporelle	128
6.3.2	Comparaison de trois modèles de constante diélectrique (modèles A, D et E)	131
6.3.2.1	Variation spatiale	131
6.3.2.2	Variation temporelle	132
6.3.3	Optimisation d'autres paramètres : albédo et rugosité (modèles F, G et H)	133
6.3.3.1	Modèle F : albédo optimisé et rugosité nulle	133
6.3.3.2	Modèle G : rugosité optimisée et albédo nul	139
6.3.3.3	Modèle H : albédo optimisé et dépendance de la rugosité à l'hu- midité de surface	139
6.3.4	Comparaison des épaisseurs optiques de la végétation	144

6.4 Conclusions	147
7 Conclusions générales	149
A Intervalle de confiance	155
A.1 Intervalle de confiance pour une population à distribution normale	155
A.2 Intervalle de confiance pour un coefficient de corrélation	155
A.3 Intervalle de confiance pour un écart-type	156
B Arbres de régression pour chaque continent	157
C Modèles de constante diélectrique du sol	163
C.1 Modèle de Wang-Schmugge (Wang & Schmugge, 1980)	163
C.2 Modèle empirique de Dobson (Dobson et al., 1985) avec modification de (Pe- plinski et al., 1995) pour la bande L	164
C.3 Modèle semi-empirique de Mironov (Mironov et al., 2004)	165
D Modélisations de la température efficace	167
D.1 Théorie	167
D.2 Modèle de Choudhury	167
D.3 Modèle de Wigneron	168
D.4 Modèle de Holmes	168
E Retrouver l'épaisseur optique de la végétation à partir des températures de brillance	169
F Étude du paramètre d'albédo optimisé avec une rugosité dépendante ou non de l'humidité du sol	171
F.1 Albédo optimisé	171
F.2 Albédo optimisé avec la rugosité dépendante de l'humidité du sol	174
G Historique des versions du produit d'humidité SMOS de niveau 2	179

Table des figures

1.1	Impact du changement climatique sur les ressources en eau d'un bassin versant, source web : Ressources Naturelles Canada (2008).	3
1.2	Schéma d'une plante qui se nourrit en eau grâce à ses racines et l'eau autour des particules du sol, source web : Levaray (2004).	4
1.3	Schéma simplifié du cycle de l'eau, source web : Stéphanoise des Eaux (2007). . .	5
1.4	Theta Probe : instrument permettant de calculer l'humidité du sol à travers la mesure de la constante diélectrique du sol.	6
1.5	Spectre électromagnétique, source web : NASA (2008).	7
1.6	Organisation de la thèse et du manuscrit.	10
2.1	Onde électromagnétique avec son champ électrique E et son champ magnétique B, source web : Bioinformatics (2002).	12
2.2	Définition des polarisations verticales et horizontales : V dans le plan perpendiculaire au plan d'incidence et H dans le plan parallèle à la surface de la Terre (ATBD, 2010).	13
2.3	Sensibilité de la température de brillance à la biomasse, à l'eau liquide dans les nuages, à l'eau dans l'atmosphère, à l'humidité du sol et à la rugosité de surface en fonction de la fréquence (Kerr, 1996).	16
2.4	Le satellite SMOS lors du lancement à Plesetsk le 2 Novembre 2009 (a) avec une représentation du satellite en orbite (b) ainsi qu'une vue détaillée d'un receveur (c) qui permet de capter les ondes électromagnétiques émises par la Terre. Images ESA.	17
2.5	Données SMOS de niveau 1 organisées par images acquises par le radiomètre appelées snapshots. Un point (indiqué en rouge) est donc vu à plusieurs reprises par les capteurs à différents angles d'incidence.	18
2.6	Données SMOS de niveau 1c : températures de brillance TB dans le référentielles de l'antenne (a - polarisations X et Y) auxquelles il faut appliquer une rotation pour les obtenir dans le référentiel terrestre (b - polarisations Horizontale et Verticale).	19
2.7	Modèle de transfert radiatif avec les différentes contributions de l'atmosphère, de la végétation et du sol. Le radiomètre SMOS observe la somme de ces contributions et afin de retrouver l'humidité du sol, il faut pouvoir estimer certaines de ces contributions.	20

2.8	Différences entre les considérations de SMOS et des satellites comme ASMR-E concernant leurs champs de vue : SMOS prend en compte l'hétérogénéité de la surface en fractionnant le champ de vue et y applique une pondération correspondant au diagramme d'antenne alors que les satellites comme AMSR-E considèrent une surface homogène.	21
2.9	Exemple d'une zone de travail représentée (a) à travers 30 classes ECOCLIMAP (il en existe 256 en tout) à une résolution de 30" d'arc; (b) les classes ECOCLIMAP ont été regroupées en 8 classes génériques : basse végétation (FNO), forêt (FFO), zone humide (FWL), eau pure (FWP), eau salée (FWS), zone aride (FEB), glace (FEI) et zone urbaine (FEU); (c) les 8 classes génériques à une résolution de 4km.	23
3.1	Le satellite Aqua.	27
3.2	L'instrument ASCAT.	27
3.3	Les quatre bassins versants tests pour comparer les données SMOS aux Etats-Unis. Pour chaque bassin versant, une représentation de l'occupation du sol est fournie par ECOCLIMAP (FTM : topographie modérée, FTS : topographie forte, FEU : zone urbaine, FEI : glace, FEB : zone aride, FWS : eau salée, FWP : eau douce, FWL : zone humide, FFO : forêt et FNO : zone de végétation basse) . . .	29
3.4	Walnut Gulch, (Jackson et al., 2010).	30
3.5	Little Washita, (Jackson et al., 2010).	31
3.6	Little River, (Jackson et al., 2010).	31
3.7	Reynolds Creek, (Jackson et al., 2010).	31
3.8	Série temporelle de l'humidité du sol (<i>SM</i>) du matin sur Walnut Gulch retrouvée par SMOS, ECMWF, VUA et les données terrain (1er panneau); humidité du sol linéairement normalisée d'ASCAT (<i>rescaled SM</i>) et les données terrain (2ème panneau); les données de végétation LAI d'ECOCLIMAP ainsi que l'épaisseur de neige d'ECMWF avec la température mesurée au sol (<i>snow depth</i> et <i>soil temperature</i>). Les jours de neige et de gel sont indiqués en ombres bleues sur le panneau du bas.	36
3.9	Série temporelle de l'humidité du sol (<i>SM</i>) du matin sur Little Washita vue par SMOS, ECMWF, VUA et les données terrain (1er panneau); humidité du sol linéairement normalisés d'ASCAT (<i>rescaled SM</i>) et les données terrain (2ème panneau); les données de végétation LAI d'ECOCLIMAP ainsi que l'épaisseur de neige d'ECMWF avec la température mesurée au sol (<i>snow depth</i> et <i>soil temperature</i>). Les jours de neige et de gel sont indiqués en ombres bleues sur le panneau du bas.	37
3.10	Série temporelle de l'humidité du sol (<i>SM</i>) du matin sur Little River vue par SMOS, ECMWF, VUA et les données terrain (1er panneau); humidité du sol linéairement normalisés d'ASCAT (<i>rescaled SM</i>) et les données terrain (2ème panneau); les données de végétation LAI d'ECOCLIMAP ainsi que l'épaisseur de neige d'ECMWF avec la température mesurée au sol (<i>snow depth</i> et <i>soil temperature</i>). Les jours de neige et de gel sont indiqués en ombres bleues sur le panneau du bas.	39

3.11	Série temporelle de l'humidité du sol (SM) du matin sur Reynolds Creek vue par SMOS, ECMWF, VUA et les données terrain (1er panneau); humidité du sol linéairement normalisés d'ASCAT (<i>rescaled SM</i>) et les données terrain (2ème panneau); les données de végétation LAI d'ECOCLIMAP ainsi que l'épaisseur de neige d'ECMWF avec la température mesurée au sol (<i>snow depth</i> et <i>soil temperature</i>). Les jours de neige et de gel sont indiqués en ombres bleues sur le panneau du bas.	41
3.12	Comparaison des humidités des sols retrouvées par SMOS, ECMWF, VUA et ASCAT normalisé avec les données d'humidité mesurées sur les quatre bassins versants en 2010 (matin uniquement). (*) dénote un résultat statistique non significatif avec une valeur p supérieure à 0.05.	42
3.13	Statistiques sur les comparaisons entre les produits d'humidité des sols de SMOS, ECMWF, VUA et les produits ASCAT normalisés linéairement avec les humidités mesurées sur les quatre bassins versants tests. Les 4 sous figures correspondent au coefficient de corrélation R (a), l'erreur d'estimation SEE (b), le biais (c) et l'erreur globale RMSE (d). Les données du matin (AM) et de l'après-midi (PM) ont été traitées séparément et les intervalles de confiance à 95% sont indiqués pour R et SEE.	43
3.14	Distributions cumulées des jeux de données ASCAT et des mesures terrain sur les quatre bassins versants pour 2010.	44
3.15	Série temporelle de l'humidité du sol d'ASCAT normalisée statistiquement avec les mesures terrain sur les quatre bassins versants en 2010.	45
3.16	Comparaison entre le biais du produit VUA et la végétation (LAI) et la température de surface (T_s). Le biais est anti-corrélé au LAI et à la température du sol. Aucune régression n'a été réalisée sur Reynolds Creek car elle n'était pas statistiquement significative. R est le coefficient de corrélation, [.] est son intervalle de confiance à 95% et N est le nombre de points.	47
3.17	Températures de brillance AMSR-E (T_b 6 GHz et 36 GHz), rapport de polarisation (RP) à 6 GHz (différence des polarisations divisée par leur somme), humidité des sols (sm) retrouvée par VUA et mesurée sur le site de Little Washita, épaisseur optique de la végétation (tau) retrouvée par VUA et LAI d'ECOCLIMAP, température du sol (T) retrouvée par VUA et mesurée sur le site.	48
3.18	Tests de sensibilité de l'algorithme VUA sur l'effet de la végétation (a) et de la température du sol (b).	49
3.19	Série temporelle en 2010 de l'épaisseur optique de la végétation retrouvée par SMOS puis lissée sur 20 jours et accompagnée du LAI d'ECOCLIMAP afin de comparer l'évolution de la végétation sur Walnut Gulch et Little Washita.	50
3.20	Série temporelle en 2010 de l'épaisseur optique de la végétation retrouvée par SMOS puis lissée sur 20 jours et accompagnée du LAI d'ECOCLIMAP afin de comparer l'évolution de la végétation sur Little River et Reynolds Creek.	51
4.1	Humidité des sols (m^3/m^3) retrouvée par SMOS le 15 Septembre 2010 à 6 :00h, heure solaire locale.	57
4.2	Humidité des sols (m^3/m^3) retrouvée par NSIDC à partir des observations d'AMSR-E le 15 Septembre 2010 à 1 :30h, heure solaire locale.	58
4.3	Humidité des sols (m^3/m^3) retrouvée par VUA à partir des observations d'AMSR-E le 15 Septembre 2010 à 1 :30h, heure solaire locale.	59
4.4	Indice d'humidité des sols retrouvé par ASCAT le 15 Septembre 2010 à 9 :30h, heure solaire locale.	60

4.5	Humidité des sols (m^3/m^3) retrouvée par le modèle ECMWF le 15 Septembre 2010 à 6 :00h, heure solaire locale.	61
4.6	La grille de SMOS est une grille hexagonale où tous les points sont espacés de 15 km (a). Exemple de l'Australie en 2010 avec les dates communes entre SMOS, VUA et ECMWF (b), puis en considérant les 6 plus proches voisins dans un rayon maximum de 30 km (c), et des 18 plus proches voisins dans un rayon maximum de 60 km.	63
4.7	Erreurs relatives de SMOS, VUA et ECMWF présentées à l'échelle globale et issues de la triple collocation appliquée en 2010 sur SMOS-VUA-ECMWF. Le meilleur jeu de données est celui dont l'erreur est significativement la plus basse.	65
4.8	Biais et facteurs d'échelle de VUA et d'ECMWF par rapport à SMOS, calculés à partir des équations (4.6) et (4.7).	66
4.9	Erreurs relatives de SMOS, NSDIC et ASCAT présentées à l'échelle globale et issues de la triple collocation appliquée en 2010 sur SMOS-NSIDC-ASCAT. Le meilleur jeu de données est celui dont l'erreur est significativement la plus basse.	68
4.10	Biais et facteurs d'échelle de NSIDC et d'ASCAT par rapport à SMOS, calculés à partir des équations (4.6) et (4.7).	69
4.11	Erreurs relatives de SMOS, VUA et ASCAT présentées à l'échelle globale et issues de la triple collocation appliquée en 2010 sur SMOS-VUA-ASCAT. Le meilleur jeu de données est celui dont l'erreur est significativement la plus basse.	71
4.12	Les biais et facteurs d'échelle de VUA et d'ASCAT par rapport à SMOS. Ils ont été calculés à partir des équations (4.6) et (4.7).	72
4.13	Médiane (trait rouge), 1er et 3ème quartile (boîte bleue) de chaque paramètre étudié à l'échelle globale et continentale en 2010 (à l'exception des fraction d'eau salée et de glace car elles étaient nulles ou très proches de 0 pour tous les continents). L'erreur SMOS est issue de la triple collocation entre SMOS, VUA et ECMWF.	74
4.14	Influence de chaque paramètre sur l'erreur SMOS à l'échelle globale et continentale : pourcentage de sable et d'argile (% sand, % clay), probabilité moyenne d'occurrence de RFI en 2010 (RFI), fraction de forêt (FFO), fraction de zone humide (FWL), fraction d'eau pure (FWP), fraction d'eau salée (FWS), fraction de sol nu ou aride (FEB) et la fraction de zone urbaine (FEU). Une part de diagramme détachée représente une influence négative et une part attachée une influence positive.	77
4.15	Arbre de régression de l'erreur SMOS à l'échelle globale. Dans les rectangles sont inscrites les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre et/ou le pourcentage de points qui sont concernés. Si la condition est remplie, alors la branche <i>vrai</i> de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.	80

5.1	Représentation temporelle de quelques satellites qui ont précédé ou qui vont suivre la mission SMOS pour retrouver l'humidité des sols à l'échelle globale : SMMR (<i>Scanning Multichannel Microwave Radiometer</i> , 1978-1987), SSM/I (<i>Special Sensor Microwave/Imager</i> , 1987-, satellites F8 à F15), AMSR-E (<i>Advanced Microwave Scanning Radiometer-Earth observation system</i> , 2002-2011), ERS-ASCAT (<i>Advanced Scatterometer</i> , 2006-), SMOS (<i>Soil Moisture and Ocean Salinity</i> , 2009-), Aquarius (2011-), SMAP (<i>Soil Moisture Active Passive</i> , prévu en 2014). Les bandes de fréquence et les heures de passage sont également indiquées pour chaque mission spatiale.	87
5.2	Série temporelle de l'humidité des sols dans le désert du Namib en utilisant les observations des satellites SMMR, SSM/I, AMSR-E, ASCAT et SMOS (en m^3/m^3).	88
5.3	Principe du CDF matching où l'égalité des probabilités est imposée. Pour une valeur x donnée, trouver la valeur y telles que les fonctions de répartition soient égales : $G_Y(y) = F_X(x)$ ou $y = G_Y^{-1}(F_X(x))$	90
5.4	Exemple d'application du CDF (<i>Cumulative Density Function</i>) matching. La série temporelle bleue est ramenée au niveau de la série rouge en faisant correspondre leur distributions statistiques.	91
5.5	Exemple de comparaison des fonctions de répartition (CDF) de SMOS et d'AMSR-E/VUA où les probabilités sont très dispersées et où l'hypothèse de les égaliser avec un CDF matching ferait perdre de l'information.	92
5.6	Représentations des neuf copules testées dans cette étude montrant leurs caractéristiques sous forme de nuages de points (abscisses : CDF de la 1ère variable ; ordonnées : CDF de la 2ème variable).	95
5.7	Exemple d'application de la méthode des copules. La série temporelle bleue est ramenée au niveau de la série rouge en faisant correspondre leur distributions statistiques.	96
5.8	Différences de simulations des humidités (SM en m^3/m^3) entre le CDF matching et la méthode des copules. Les données d'origine sont en points bleus, les données simulées par CDF matching en vert et celles simulées par copules en rouge. L'incertitude de la méthode des copules est indiquée en ombre rouge. Chaque ligne correspond à un site test et chaque colonne à une saison de l'année.	98
5.9	Simulations pour l'année 2010 : SMOS, VUA, humidités simulées par CDF matching et copules et les mesures terrain sur les quatre bassins versants. Puisque les données terrain sont des moyennes de plusieurs stations, leur écart-type est représenté en ombre noire montrant ainsi la variabilité spatiale à l'intérieur de chaque bassin versant.	100
5.10	Séries temporelles homogènes des humidités des sols (SM en m^3/m^3) simulées par CDF matching (vert) et par copules (violet) de 2003 à 2010. Les mesures terrain sur les 4 bassins versants sont indiquées en noir alors que les données VUA d'origine sont en gris. L'hiver ne faisant pas partie de l'étude, aucune simulation n'a été réalisée sur cette saison.	103
5.11	Choix de la famille de copules la plus adaptée (a) pour faire un lien entre les humidités SMOS et VUA sur les zones de prairies pour les mois d'Avril, Mai et Juin. La méthode de Huard fournit également un coefficient de poids (b) permettant de juger ce choix : 1 pour une certitude, 0 quand le choix est très incertain.	105

5.12	Choix de la famille de copules la plus adaptée (a) pour faire un lien entre les humidités SMOS et VUA sur les zones de cultures pour les mois d'Octobre, Novembre et Décembre. La méthode de Huard fournit également un coefficient de poids (b) permettant de juger ce choix : 1 pour une certitude, 0 que le choix n'est pas sûr du tout.	106
6.1	Schéma des contributions au signal observé par le radiomètre en orbite avec trois couches : le sol (émissivité e_g , réflectivité r_g , température effective du sol T_g), la végétation (épaisseur optique τ , température T_c , albedo ω) et l'atmosphère (épaisseur optique de l'atmosphère τ , émission de l'atmosphère T_{Batm} , émission cosmique T_{Bsky}).	111
6.2	Influence de l'humidité sur le profil angulaire des températures de brillance simulées en bande L (modèle LPRM, Owe et al. (2001)). L'humidité a tendance à diminuer les températures de brillance mais cette différence s'atténue avec l'angle d'observation.	113
6.3	Différence de modélisation de la constante diélectrique pour un sol d'humidité $0.15 \text{ m}^3/\text{m}^3$ composé de 10% de sable et 50% d'argile en utilisant les modèles de Wang-Schmugge, Dobson et Mironov.	114
6.4	Influence de la rugosité et du paramètre de mélange de polarisation sur le profil angulaire de la température de brillance simulée en bande L (modèle LPRM, Owe et al. (2001)). Une augmentation de la rugosité aura tendance à augmenter la température de brillance. Le paramètre de mélange de polarisations rapproche les polarisations verticales et horizontales de la température de brillance.	115
6.5	Influence de la végétation sur le profil angulaire de la température de brillance simulée en bande L (modèle LPRM, Owe et al. (2001)). Plus l'épaisseur optique τ est élevée, plus la température de brillance au nadir est élevée et plus la différence de polarisation diminue. L'albédo abaisse le niveau moyen des températures de brillance, comme une translation verticale de tout le profil.	116
6.6	Schéma de l'algorithme LPRM (<i>Land Parameter Retrieval Model</i>) en 5 phases : calcul de la constante diélectrique, calcul de la réflectivité du sol, prise en compte de la rugosité du sol, influence de la végétation puis modélisation du transfert radiatif afin de modéliser une température de brillance qui est ensuite comparée à la température acquise par satellite. L'humidité du sol retrouvée est celle qui minimise l'écart entre ces deux températures de brillance. Les flèches verticales représentent les entrées alors que les flèches verticales des paramètres.	118
6.7	Comportement du modèle LPRM sur un profil angulaire de température de brillance <i>propre</i> . La colonne de gauche (a,c,e) correspond au modèle original, la colonne de droite (b,d,f) au modèle avec le rapport de polarisation lissé. Le panneau du bas (g) correspond au modèle itératif. Les TB observées SMOS ainsi que les TB modélisées sont en (a,b,g), le rapport de polarisation par rapport à l'angle en (c,d) et l'épaisseur optique de la végétation pour chaque angle d'observation en (e,f).	123
6.8	Comportement du modèle LPRM sur un profil angulaire de température de brillance bruité. La colonne de gauche (a,c,e) correspond au modèle original, la colonne de droite (b,d,f) au modèle avec le rapport de polarisation lissé. Le panneau du bas (g) correspond au modèle itératif. Les TB observées SMOS ainsi que les TB modélisées sont en (a,b,g), le rapport de polarisation par rapport à l'angle en (c,d) et l'épaisseur optique de la végétation pour chaque angle d'observation en (e,f).	124

6.9	Représentations spatiales des humidités retrouvées avec les trois approches LPRM avec celles de SMOS sur les Etats-Unis (16/09/2010) et sur l'Australie(16/09/2010) : (a,e) LPRM basique, (b,f) LPRM avec le rapport de polarisation (RP) lissé, (c,g) LPRM itératif et (d,h) SMOS.	126
6.10	Comparaison des trois approches d'application du modèle LPRM avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis et en Australie : (a,d) LPRM basique, (b,e) LPRM avec le rapport de polarisation (RP) lissé et (c,f) LPRM itératif.	127
6.11	Représentations temporelles des humidités retrouvées avec les trois approches du modèle LPRM sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) LPRM basique, (c,d) LPRM avec rapport de polarisation (RP) lissé et (e,f) LPRM itératif.	128
6.12	Comparaison des humidités retrouvées par les trois approches du modèle LPRM avec les humidités mesurées sur Little Washita (LW) et Walnut Gulch (WG) : (a,e) LPRM basique, (b,f) LPRM avec rapport de polarisation (RP) lissé et (c,g) LPRM itératif. Les statistiques sont présentées en (d,h).	129
6.13	Représentations spatiales des humidités retrouvées à partir des trois modèles de constante diélectrique avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,e) LPRM avec le modèle de Wang-Schmugge, (b,f) LPRM avec le modèle de Dobson, (c,g) LPRM avec le modèle de Mironov et (d,h) SMOS avec le modèle de Dobson.	130
6.14	Comparaison des humidités retrouvées avec le LPRM et les trois modèles de constante diélectrique avec celles de SMOS sur les Etats-Unis et sur l'Australie : (a,d) modèle de Wang-Schmugge, (b,e) modèle de Dobson et (c,f) modèle de Mironov.	131
6.15	Représentations temporelles des humidités retrouvées avec le LPRM et trois modèles de constante diélectrique sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.	132
6.16	Comparaison des humidités retrouvées par le modèle LPRM et trois modèles de constante diélectrique sur Little Washita (LW) et Walnut Gulch (WG) : (a,e) Wang-Schmugge, (b,f) Dobson et (c,g) Mironov. Les statistiques sont présentées en (d,h).	133
6.17	Représentations spatiales des humidités retrouvées à partir des trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,e) LPRM avec le modèle de Wang-Schmugge, (b,f) LPRM avec le modèle de Dobson, (c,g) LPRM avec le modèle de Mironov et (d,h) SMOS avec le modèle de Dobson.	134
6.18	Comparaison des humidités retrouvées avec le LPRM et les trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec celles de SMOS sur les Etats-Unis et sur l'Australie : (a,d) modèle de Wang-Schmugge, (b,e) modèle de Dobson et (c,f) modèle de Mironov.	136
6.19	Représentations temporelles des humidités retrouvées avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.	137

6.20	Comparaison des humidités retrouvées par le modèle LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω sur Little Washita (LW) et Walnut Gulch (WG) : (a,e) Wang-Schmugge, (b,f) Dobson et (c,g) Mironov. Les statistiques sont présentées en (d,h).	138
6.21	Représentations spatiales des humidités retrouvées à partir des trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,e) LPRM avec le modèle de Wang-Schmugge, (b,f) LPRM avec le modèle de Dobson, (c,g) LPRM avec le modèle de Mironov et (d,h) SMOS avec le modèle de Dobson.	140
6.22	Comparaison des humidités retrouvées avec le LPRM et les trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol avec celles de SMOS sur les États-Unis et sur l'Australie : (a,d) modèle de Wang-Schmugge, (b,e) modèle de Dobson et (c,f) modèle de Mironov.	141
6.23	Représentations temporelles des humidités retrouvées avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.	142
6.24	Comparaison des humidités retrouvées par le modèle LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol sur Little Washita (LW) et Walnut Gulch (WG) : (a,e) Wang-Schmugge, (b,f) Dobson et (c,g) Mironov. Les statistiques sont présentées en (d,h).	143
6.25	Représentations spatiales des épaisseurs optiques de la végétation retrouvées.	145
6.26	Représentations spatiales des épaisseurs optiques de la végétation retrouvées.	146
B.1	Arbre de régression de l'erreur SMOS à l'échelle continentale. Dans les rectangles sont inscrits les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre de points qui sont concernés. Si la condition est remplie, alors la branche <i>vraie</i> de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.	158
B.1	Arbre de régression de l'erreur SMOS à l'échelle continentale. Dans les rectangles sont inscrits les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre de points qui sont concernés. Si la condition est remplie, alors la branche <i>vraie</i> de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.	159
B.1	Arbre de régression de l'erreur SMOS à l'échelle continentale. Dans les rectangles sont inscrits les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre de points qui sont concernés. Si la condition est remplie, alors la branche <i>vraie</i> de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.	160

B.1	Arbre de régression de l'erreur SMOS à l'échelle continentale. Dans les rectangles sont inscrits les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre de points qui sont concernés. Si la condition est remplie, alors la branche <i>vraie</i> de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.	161
F.1	Représentations spatiales de l'albédo ω optimisée avec les trois modèles de constante diélectrique en libérant le paramètre d'albédo ω sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,b) LPRM avec le modèle de Wang-Schmugge, (c,d) LPRM avec le modèle de Dobson, (e,f) LPRM avec le modèle de Mironov.	172
F.2	Représentations temporelles des albédos retrouvés avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.	173
F.3	Représentations spatiales de l'albédo ω retrouvé à partir des trois modèles de constante diélectrique en libérant ω avec une dépendance de la rugosité à l'humidité du sol avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,b) LPRM avec le modèle de Wang-Schmugge, (c,d) LPRM avec le modèle de Dobson, (e,f) LPRM avec le modèle de Mironov.	174
F.4	Représentations spatiales de la rugosité à 42.5° retrouvée à partir des trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,b) LPRM avec le modèle de Wang-Schmugge, (c,d) LPRM avec le modèle de Dobson, (e,f) LPRM avec le modèle de Mironov.	175
F.5	Représentations temporelles des albédos retrouvés avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.	176
F.6	Représentations temporelles des rugosités à 42.5° retrouvées avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.	177

Liste des tableaux

3.1	Caractéristiques des quatre bassins versants : le nombre de stations qui mesurent l'humidité, le climat, les précipitations annuelles, la topographie et le type majoritaire d'utilisation du sol (Jackson et al., 2010).	30
3.2	Résultats statistiques entre les produits d'humidité (satellites et modèles) et les mesures <i>in situ</i> sur les 4 bassins versants tests avec AM correspondant aux orbites ascendantes du matin pour SMOS (6 :00), ECMWF (6 :00), AMSR-E (1 :30) et descendantes du matin pour ASCAT (9 :30). Les données ASCAT ont été normalisées linéairement dans la gamme des valeurs in situ (appelées ASCATn). Les meilleurs résultats sont indiqués en gras. (*) dénote un résultat statistique non significatif qui correspond à une valeur p supérieure à 0.05.	34
3.3	Probabilité d'occurrence de RFI (<i>Radio Frequency Interference</i>) en 2010 sur les quatre bassins versants test le matin et l'après-midi (ATBD, 2010)	43
3.4	Résultats statistiques (corrélation R et intervalles de confiance IC) pour les données ASCAT normalisées statistiquement avec les mesures terrain de 2010 (* : non significatif statistiquement, valeur $p \geq 0.05$).	44
4.1	Coefficients de corrélation de la régression linéaire multiple à l'échelle globale puis continentale.	75
4.2	Pourcentages de la variance de l'erreur SMOS expliqués par chaque variable explicative : pourcentages de sable et d'argile, probabilité moyenne de RFI sur 2010 (RFI), fraction de forêt (FFO), fraction de zone humide (FWL), fraction d'eau libre (FWP), fraction d'eau salée (FWS), fraction de sol aride (FEB), fraction de surface gelée (FEI) et fraction de zone urbaine (FEU).	76
4.3	Valeurs des paramètres qui mènent à l'erreur SMOS la plus faible et la plus élevée pendant le processus de classification. Les paramètres sont indiqués dans l'ordre d'apparition dans l'arbre de régression et les erreurs moyennes SMOS sont entre parenthèses.	81
5.1	Coefficient de corrélation (R) entre les mesures terrain d'humidité à 1 :30h et 6 :00h sur les quatre bassins versants test en 2010 (Walnut Gulch, Little Washita, Little River et Reynolds Creek) avec son intervalle de confiance à 95%. N est le nombre de dates communes disponibles en 2010.	89
5.2	Les neuf copules testées dans cette étude : définition, intervalle du paramètre θ et famille.	94

5.3	Résultats statistiques des simulations des copules et du CDF matching. Ces simulations sont comparées aux mesures terrain d'humidité en 2010 divisées en 4 saisons : Mars-Avril-Mai (MAM), Juin-Juillet-Août (JJA), Septembre-Octobre-Novembre (SON), mais il n'y avait pas assez de données disponibles pour la saison d'hiver. Les résultats significativement meilleurs sont indiqués en gras et les RMSE sont en m^3/m^3	101
5.4	Résultats statistiques de la comparaison entre les séries temporelles simulées de l'humidité de 2003 à 2010. L'historique d'origine est représenté par VUA. Les résultats significativement meilleurs sont indiqués en gras et les RMSE sont en m^3/m^3	104
6.1	Modifications apportées au modèle LPRM pour son adaptation aux acquisitions multi-angulaires en bande L de SMOS.	120
6.2	Options testées pour l'adaptation du LPRM aux acquisitions multi-angulaires en bande L de SMOS concernant le modèle de constante diélectrique, l'albédo et la rugosité (modèles A à H). RP pour rapport de polarisation, Eq. 6.16.	121

Chapitre 1

Introduction générale

Dans un contexte de changement climatique, de nombreuses études portent sur l'observation de plusieurs paramètres physiques tels que la température du sol, la végétation ou le contenu en eau du sol. L'humidité du sol fait partie des traceurs observés pour surveiller l'évolution du cycle hydrologique sur Terre. De nombreux travaux prédisent une augmentation de la température de la surface terrestre mais il n'y a aucune certitude concernant un possible ralentissement ou accélération du cycle de l'eau. Le besoin d'observer pour mieux prédire se fait de plus en plus vital pour une meilleure gestion des ressources naturelles.

Les satellites d'observation de la Terre contribuent à répondre à ce besoin. Parmi eux, la première mission à avoir été spécialement conçue pour l'observation de l'humidité des sols à l'échelle globale est la mission SMOS (*Soil Moisture and Ocean Salinity*). Durant ma thèse, j'ai utilisé, contrôlé la qualité et validé les données d'humidité SMOS sur plusieurs sites test ainsi qu'à l'échelle globale. Le plan de ce manuscrit et de mes travaux est présenté à la fin de ce chapitre.



1.1 Contexte

1.1.1 Le changement climatique

Plus personne n'ignore la problématique du changement climatique et de nombreuses études portent sur son observation et sa prédiction, d'autant plus qu'il est fort probable que ce soit lié à l'activité humaine à travers le rejet croissant des gaz à effet de serre. Ces changements climatiques se traduisent dans un premier temps par une élévation de la température de la surface du sol, par une augmentation de la hauteur moyenne des océans ou encore par une désertification de certaines régions (IPCC - *Intergovernmental Panel on Climate Change*, 2007). Alors que l'eau est un composant essentiel de la vie sur Terre, l'impact du dérèglement climatique sur les ressources en eau de la Terre n'est pas encore précisément déterminé (fig. 1.1).

Afin d'observer ces changements climatiques, il y a un besoin évident d'une observation continue et globale de certains paramètres physiques tels que la température du sol, la végétation, la salinité des océans ou le stock d'eau douce. Pour cela, le GCOS (*Global Climate Observing System*) a répertorié 50 variables climatiques essentielles réparties en trois domaines (atmosphérique, océanique et terrestre) sur lesquels les recherches et l'observation sont primordiales pour mesurer et prédire les divers changements à l'échelle globale et régionale. L'humidité du sol, ou le contenu en eau du sol, fait partie de ces variables à surveiller pour suivre l'évolution du cycle hydrologique sur les terres émergées (World Meteorological Organization et al., 2010).

L'eau sur Terre n'est pas uniquement exposée au bouleversement climatique car se pose également la question de l'évolution de sa demande avec une augmentation constante de la population mondiale ainsi qu'un accroissement des infrastructures. Pour être en mesure de prendre les meilleures décisions politiques qui soient adaptées à une gestion durable, un suivi local et régional de la ressource en eau douce est indispensable.

Les satellites contribuent à l'observation du changement climatique, de l'évolution du cycle de l'eau et plus particulièrement de l'humidité des sols depuis plus de trois décennies maintenant.

1.1.2 L'eau sur Terre

1.1.2.1 Quelle quantité d'eau y a-t-il sur Terre ?

La surface terrestre est majoritairement composée d'eau, environ 70% de sa surface en est recouverte. L'eau existe également dans l'air sous forme de vapeur, d'eau liquide et de glace ainsi que dans le sol en tant qu'eau souterraine. La réserve totale en eau sur la planète est de 1.400.000.000 km³. Tous les ans, 119.000 km³ d'eau tombent sur les terres sous forme de précipitations alors que 74.200 km³ d'eau s'évaporent dans l'atmosphère par le phénomène d'évapo-transpiration du sol et de la végétation. Sur les océans et les lacs, 450.000 km³ d'eau tombent chaque année et 502.800 km³ s'évaporent (Outwater, 1997).

Il est estimé à 2.200 km³ la quantité d'eau douce qui s'écoule dans les sols alors que 135.000 km³ d'eau sont stockés sous diverses formes : dans l'atmosphère sous forme de vapeur, dans les sols sous forme d'humidité, dans les lacs, les marécages, les rivières ou la végétation sous forme d'eau liquide. Les eaux souterraines et l'eau douce stockée représentent une ressource disponible en eau douce contrairement à l'eau douce stockée dans les glaciers et les calottes glaciaires situés principalement aux pôles et au Groenland qui n'est pas disponible alors qu'elle représente 69.5% de l'eau douce totale de la Terre (Outwater, 1997).

1.1.2.2 Qu'est ce qu'une quantité d'eau dans le sol ?

Le sol est un milieu qui stocke et laisse circuler l'eau. Une couche superficielle de terre typique (les quelques premiers centimètres voire une dizaine de centimètres sous la surface)

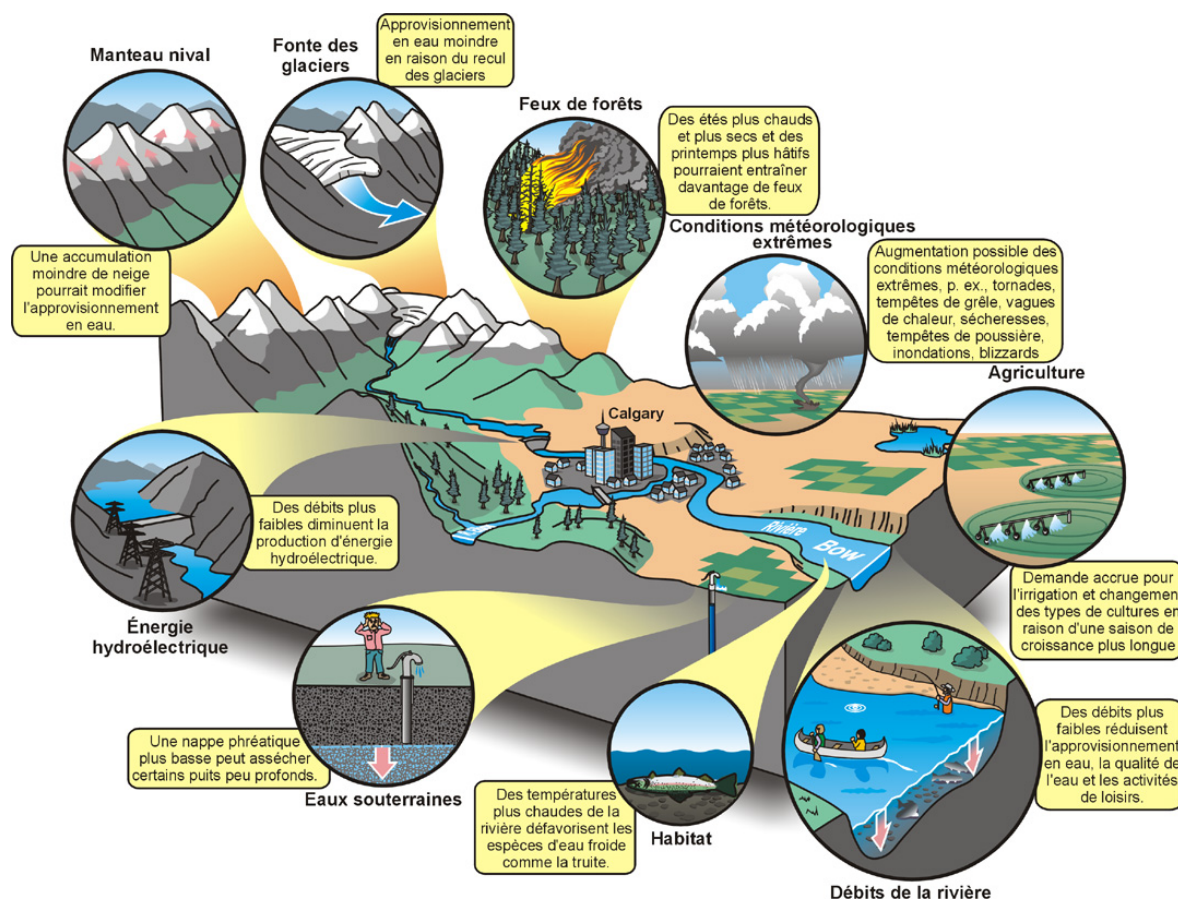


FIGURE 1.1 – Impact du changement climatique sur les ressources en eau d’un bassin versant, source web : Ressources Naturelles Canada (2008).

se décompose en plusieurs éléments : environ 45% de son volume est occupé par de la matière minérale (particules du sol), 5% par des résidus organiques (matière produite par les être vivants, végétaux ou animaux), le reste étant de l’espace occupé soit par l’eau soit par l’air. La quantité et la taille de ces espaces (ou pores) sont déterminées par la texture, la densité et la structure du sol (Scherer et al., 1996).

L’eau dans le sol est présente de deux façons : autour des particules du sol ou dans les pores. L’eau des pores peut être représentée sous deux formes : l’eau gravitationnelle et l’eau capillaire. L’eau gravitationnelle circule en général vers le bas dû à la force gravitationnelle. L’eau capillaire est la plus importante pour la croissance des plantes puisque c’est celle qui reste accrochée aux particules de terre contre la force de gravité et qui est donc susceptible d’être captée par les racines (figure 1.2). Quand l’eau s’infiltré dans le sol, les pores se remplissent. Au fur et à mesure que les pores se remplissent, l’eau se déplace par gravité et capillarité. L’eau continue de se déplacer vers le bas jusqu’à ce qu’un équilibre soit atteint entre les forces gravitationnelles et capillaires (Scherer et al., 1996).

Le terme *humidité du sol* correspond au contenu en eau présent dans le sol. Il est exprimé la plupart du temps de façon volumétrique (m^3/m^3) en prenant le rapport entre le volume occupé par l’eau et le volume total. Ses valeurs vont de 0 (pas d’eau dans le sol - sol très aride) jusqu’à la porosité qui est une valeur spécifique à chaque sol. La porosité est la fraction de volume occupée par les pores. Une humidité volumétrique ne peut dépasser la porosité puisqu’elle ne peut occuper plus de volume que le volume libre. On parle alors de saturation du sol qui peut

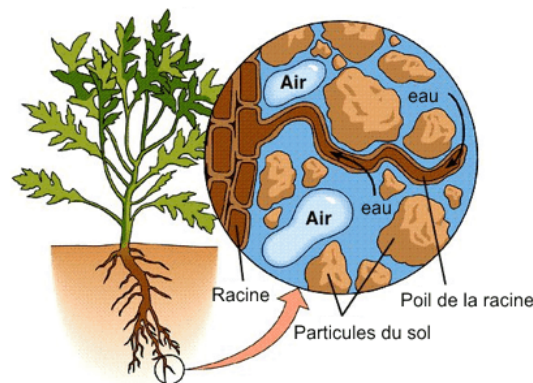


FIGURE 1.2 – Schéma d'une plante qui se nourrit en eau grâce à ses racines et l'eau autour des particules du sol, source web : Levaray (2004).

être atteint lors de fortes pluies notamment sur les premiers centimètres de terre sous la surface.

1.1.3 Importance de l'humidité du sol dans le cycle global de l'eau

1.1.3.1 Le cycle de l'eau

L'eau est présente sous différentes formes sur Terre. L'énergie solaire est le moteur de la transformation et du transport de cette eau à travers tous ses états. Le cycle de l'eau est le parcours de l'eau sur les terres, dans les océans et dans l'atmosphère (figure 1.3). L'eau est d'abord naturellement présente dans les océans. Sous l'effet du soleil, l'eau présente à la surface s'évapore et peut former par la suite de minuscules gouttelettes de glace puis des nuages. Avec le vent, ces nuages se déplacent au dessus des terres. Soumise à des processus de condensation et d'agrégation à l'intérieur de ces nuages, les gouttelettes deviennent trop lourdes pour demeurer en suspension dans l'atmosphère et redescendent au sol par précipitation. Les grandes surfaces d'eau comme les mers, les océans, les lacs ne sont pas les seules sources d'eau. Il existe aussi le phénomène d'évapo-transpiration des sols et des végétaux. Le sol et les plantes sont composés en partie d'eau en surface mais aussi à l'intérieur. L'eau présente en surface est soumise à évaporation et l'eau interne peut également être transpirée.

L'eau ainsi venue du ciel est ensuite infiltrée dans le sol. Lors de cette infiltration, une certaine quantité est stockée autour des particules du sol, une autre partie ruisselle à la surface du sol et le reste se déplace vers des zones plus profondes jusqu'à atteindre une nappe phréatique qui peut déboucher par la suite à la création d'un cours d'eau. Ces cours d'eau reviennent ensuite à la mer et le cycle peut ensuite se répéter.

1.1.3.2 Le rôle de l'humidité du sol dans le cycle de l'eau

L'humidité du sol ne représente qu'une infime partie de l'eau dans le cycle de l'eau. Ici nous désignons l'humidité par le contenu en eau du sol sur les quelques centimètres ou mètres tout au plus pendant la phase d'infiltration. Cependant cette phase reste cruciale pour la croissance des plantes. Si la mesure de cette infiltration est correcte alors on est en mesure de calculer le volume d'eau disponible pour la croissance de la végétation ou des cultures. L'enjeu le plus important est de pouvoir quantifier l'eau disponible pour satisfaire la croissance des différentes cultures et de pouvoir ainsi la maîtriser. L'irrigation dépend essentiellement de cette quantité d'eau disponible dans les sols. Si les quantités sont insuffisantes alors il faut maîtriser l'irrigation.

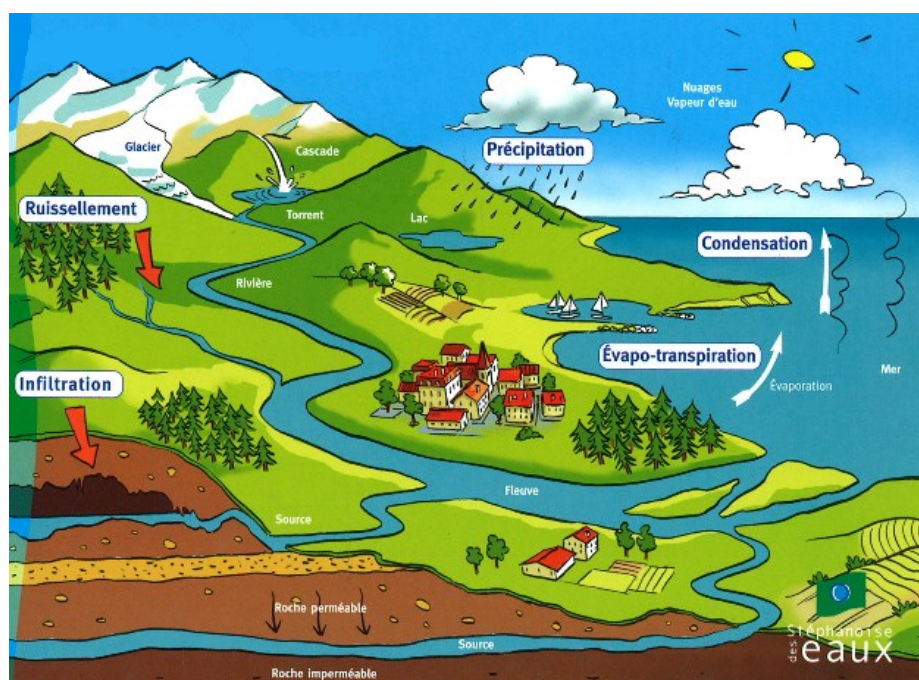


FIGURE 1.3 – Schéma simplifié du cycle de l'eau, source web : Stéphanoise des Eaux (2007).

1.1.3.3 Mesurer l'humidité du sol

Mesurer le contenu en eau du sol peut être réalisé de plusieurs façons : par prélèvement ou par mesure instrumentale.

La mesure par prélèvement consiste à prélever un volume connu de sol qui est de suite pesé puis séché dans un four spécialisé et enfin pesé une seconde fois. La différence de masse correspond à l'eau évaporée et contenue initialement dans le volume de terre prélevé. Connaissant la densité volumique de l'eau il est facile de retrouver le volume initial d'eau. Le contenu en eau du prélèvement est alors simplement le rapport entre ces deux volumes.

La mesure du contenu en eau peut également être réalisée à l'aide de sondes (capacitives ou résistives ou utilisant la mesure de réflectométrie dans le domaine temporel (TDR), (Cosh et al., 2005, Evett et al., 2006). On peut ainsi utiliser un instrument comme la Theta Probe (figure 1.4). Cet instrument mesure en réalité un champ électrique. La relation entre ce champ mesuré et l'humidité du sol est quasiment linéaire même s'il dépend fortement de la composition du sol. Une droite d'étalonnage est donc réalisée pour chaque type de sol afin de faire correspondre chaque mesure (en mV) à une valeur d'humidité.

Ces deux méthodes de mesure restent manuelles et doivent être réalisées par un opérateur. L'instrument peut cependant être relié à une centrale d'acquisition, ce qui permet de mesurer en continu l'humidité du sol. Ces stations sont très utiles pour observer le cycle journalier de cette variable. Souvent ce cycle est très prononcé et les mesures d'humidité varient fortement temporellement ainsi que spatialement. En effet il n'est pas rare d'observer de grandes différences dans les mesures dans un rayon d'un mètre, ce qui rend sa mesure d'autant plus complexe et pose la question de représentativité des mesures *in situ*. De nombreuses campagnes de mesures sont organisées pour comprendre ces différences à l'aide de prélèvements et d'instruments.

Pour une étude à l'échelle du globe, ces méthodes se révèlent laborieuses et impossibles à mettre en œuvre. Il faudrait beaucoup de temps et d'équipements pour mener à bien ce type d'étude. C'est pour cela que les satellites d'observation de la Terre sont indispensables pour observer et étudier les variables environnementales à l'échelle globale.



FIGURE 1.4 – Theta Probe : instrument permettant de calculer l’humidité du sol à travers la mesure de la constante diélectrique du sol.

1.1.4 Les satellites à l’échelle globale

De nombreux satellites ont été lancés depuis 1960 pour observer la Terre. Le premier satellite opérationnel se nommait TIROS-1 (Television InfraRed Observation Satellite lancé par la NASA) avec comme objectif l’observation de la couverture nuageuse. Malgré sa courte durée de vie (78 jours), de nombreuses photos et vidéos de la Terre ont pu être transmises aux météorologues américains afin d’améliorer leurs prédictions. La démonstration de la faisabilité d’un tel satellite et de l’utilité de ses images fut un succès.

Les premiers satellites dédiés à l’observation de la Terre n’utilisaient que quelques bandes de fréquence pour acquérir des informations. La gamme de fréquence observée est très large et chaque bande de fréquence correspond à une utilisation particulière. La télédétection est la mesure de l’énergie émise ou réfléchiée par tout corps, à la surface de la Terre ou dans l’atmosphère. Cette mesure est caractérisée par une longueur d’onde et une puissance. Les capteurs embarqués à bord des satellites mesurent la quantité d’énergie que leur apporte ce rayonnement dans un angle de vue instantané et à une fréquence donnée.

Les ondes qui composent le spectre électromagnétique (figure 1.5) ne peuvent être utilisées pour l’observation de la surface de la Terre que dans des gammes du spectre (fenêtres) dans lesquelles l’atmosphère n’interfère pas excessivement. Si le passage d’une onde à travers l’atmosphère est trop altéré, aucune information utilisable pour la surface ne pourra être retirée de cette acquisition. Il existe plusieurs fenêtres exploitables pour l’observation de la Terre :

- la gamme du visible : zone qui comprend la gamme des longueurs d’onde correspondant à la vision humaine ($0.4 - 0.7 \mu\text{m}$) et celle du proche infrarouge ($0.7 - 1.1 \mu\text{m}$), la source du rayonnement est la lumière réfléchiée du soleil (rayonnement solaire allant de 0.4 à $2 \mu\text{m}$), il est donc uniquement possible d’acquérir des informations en journée et lorsqu’il n’y a aucun nuage.
- le domaine de l’infrarouge thermique : zone qui s’étend de 7 à $25 \mu\text{m}$. Ce rayonnement transfère la chaleur de l’objet observé et sa longueur d’onde d’émission maximale est fonction de sa température. Ce transfert de chaleur peut être capté de jour comme de nuit car il est émis naturellement et n’a pas besoin de l’éclairement du soleil à condition qu’il n’y ait pas de nuage.
- le domaine micro-onde : zone qui correspond à une gamme de longueur d’onde de $25 \mu\text{m}$ à 300 mm .

En utilisant le domaine micro-onde du spectre électromagnétique, on observe des phénomènes qui sont complémentaires de ceux observés dans les fréquences plus élevées. En effet, les interactions micro-ondes sont en général gouvernées par des paramètres physiques différents de ceux

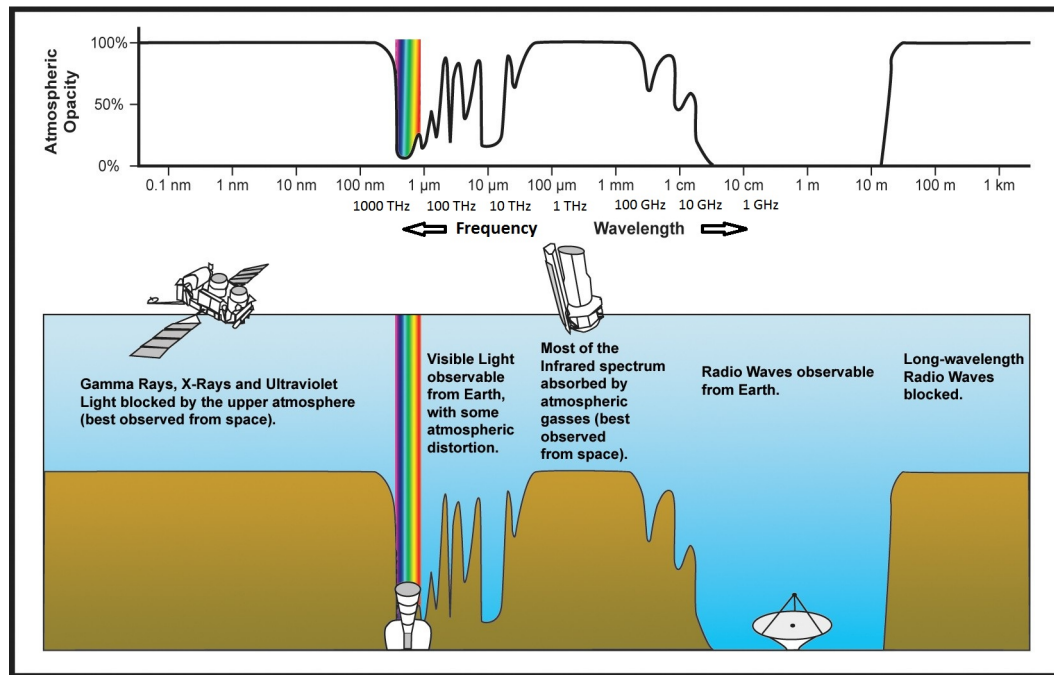


FIGURE 1.5 – Spectre électromagnétique, source web : NASA (2008).

qui sont observés dans d'autres bandes spectrales.

Par exemple, la quantité d'énergie micro-onde rayonnée par une feuille verte est liée à sa taille, sa forme et son contenu en eau plutôt que son contenu en chlorophylle ("verdure" qui est observable dans le visible). Les faibles fréquences micro-ondes ont d'autres avantages : elles ne sont pas altérées par les nuages et elles peuvent également pénétrer les premiers centimètres du sol (jusqu'à quelques mètres dans des circonstances très particulières). A contrario, les plus grandes fréquences micro-ondes permettent d'observer l'eau dans les nuages (pour les précipitations). Les micro-ondes sont naturellement émises par la nature, il n'est donc pas nécessaire de posséder une source émettrice et il est ainsi possible d'observer la Terre même de nuit, on parle alors d'observation passive. A contrario, il est possible d'observer la Terre de manière active, c'est à dire que l'instrument d'observation possède sa propre source émettrice et mesure sa réflexion, on parle alors de télédétection hyperfréquence active ou radar.

Il existe également des désagréments dans l'utilisation des micro-ondes. Le problème le plus important se situe au niveau de la technique lors de la conception de l'instrument qui va acquérir les informations. En effet, le diamètre de l'antenne est proportionnel à la longueur d'onde à résolution constante. Pour une même résolution instrumentale, le diamètre d'une antenne en micro-onde serait théoriquement 10.000 fois plus grande qu'une antenne en visible (optique) puisque la longueur d'onde passe de $1 \mu\text{m}$ à 1 cm , ce qui correspondrait à une antenne de plusieurs mètres de diamètre. Même s'il est possible de construire de telles antennes sur Terre, il est cependant techniquement difficile d'envoyer de telles structures en orbite pour observer la Terre. La technique de l'interférométrie a alors été développée et cela a permis par la suite de réduire la taille de l'instrument à envoyer en orbite et a ainsi donné naissance au concept du satellite SMOS (*Soil Moisture and Ocean Salinity*, Kerr et al. (2001)). La mission SMAP (*Soil Moisture Active Passive*, O'Neill et al. (2010)) relèvera le défi d'envoyer en orbite une structure de plusieurs mètres de diamètre avec la forme d'un parasol en 2014.

Le tableau suivant donne une correspondance entre les fréquences et les longueurs d'onde dans le domaine des micro-ondes. La gamme de fréquence a été découpée en plusieurs parties

et chaque section correspond à une lettre :

Bande	Fréquence	Longueur d'onde	Bande	Fréquence	Longueur d'onde
L	1-2 GHz	30-15 cm	Q	33-50 GHz	9.1-6 mm
S	2-4 GHz	15-7.5 cm	U	40-60 GHz	7.5-5 mm
C	4-8 GHz	7.5-3.75 cm	V	50-75 GHz	6-4 mm
X	8-12 GHz	3.75-2.5 cm	E	50-90 GHz	6-3.3 mm
Ku	12-18 GHz	2.5-1.6 cm	W	75-110 GHz	4-2.7 mm
K	18-26.5 GHz	16.6-11.3 mm	D	110-170 GHz	2.7-1.8 mm
Ka	26.5-40 GHz	11.3-7.5 mm			

Bien que de nombreux satellites ont déjà permis d'estimer l'humidité des sols depuis la fin des années 1970, SMOS est le premier satellite à avoir été spécialement conçu pour l'humidité des terres émergées, en utilisant la fréquence d'acquisition la plus adaptée à l'observation de cette variable : 1.4 GHz (Kerr et al., 2001, 2010). Il embarque un radiomètre interférométrique qui mesure la cohérence entre les 69 antennes placées sur les trois bras du satellite. Cette technique permet d'acquérir des informations à différents angles d'incidence. Cette multi-angularité permet de différencier les effets de la végétation des effets dus à l'humidité sur le signal acquis.

1.2 Objectifs de la thèse

Le premier objectif de cette thèse est de réaliser une évaluation du produit d'humidité SMOS à plusieurs échelles : locale, régionale et globale. Afin de mieux comprendre ce qu'apporte ce nouveau produit d'humidité, il sera comparé à d'autres bases de données déjà existantes. En effet, la mission SMOS apporte une nouvelle technologie et de nouveaux concepts, notamment algorithmiques, qui ont besoin d'être testés et validés.

Il existe de nombreux jeux de données d'humidité dérivés à partir d'observations de satellites, de mesures au sol ou de modèles météorologiques mais ils ne sont pas toujours utilisables ensemble en raison de leurs dynamiques et de leurs niveaux moyens d'humidité trop différents. Pour une étude climatique qui doit s'appuyer sur de longues périodes temporelles, il est essentiel de pouvoir reconstruire un historique homogène et cohérent à partir de tous ces produits d'humidité disponibles. Répondre à cette problématique constituera le deuxième objectif de cette thèse.

Enfin le troisième objectif est de proposer un modèle simple s'appuyant sur ce qui existait avant que le satellite SMOS ne soit lancé. En effet, l'algorithme permettant de retrouver l'humidité à partir des acquisitions SMOS est plus complexe que ses prédécesseurs. Dans le cadre d'une homogénéisation des produits d'humidité, un modèle simple et unique est nécessaire.

1.3 Structure de la thèse

Tout d'abord, des notions fondamentales de télédétection micro-onde sont présentées dans le chapitre suivant, accompagnées d'une description de la mission SMOS. SMOS est un concept novateur et a besoin d'être validé afin de pouvoir être utilisé dans le futur par la communauté scientifique dans des modèles hydrologiques plus larges voire dans des modèles de prédictions météorologiques.

Mon travail de thèse s'est divisé en trois majeures parties (fig. 1.6). La première phase de ce travail porte sur l'aspect de validation des données SMOS.

L'évaluation du produit SMOS a été réalisée dans un premier temps à une échelle locale où les données ont été comparées à des mesures d'humidité effectuées sur quatre bassins versants

aux États-Unis (chapitre 3). Dans un deuxième temps, les données SMOS ont été comparées à d'autres données satellites et données de modèles à l'échelle mondiale afin d'identifier les zones d'accord et de désaccord entre tous ces jeux de données (chapitre 4). Cette étude a également permis d'identifier les principales sources d'erreurs à l'échelle globale et continentale.

Après cette phase d'évaluation, les données SMOS ont été replacées dans un contexte historique. La deuxième grande partie de la thèse présente ainsi deux méthodes statistiques (le CDF matching et les copules) pour la construction de longues séries temporelles homogènes d'humidité des sols (chapitre 5).

Enfin la troisième partie reviendra sur l'algorithme de SMOS et sur ses innovations par rapport à ceux déjà existants. Pour cela, un algorithme plus simple (faisant plusieurs hypothèses de départ et ne prenant pas en compte l'hétérogénéité de la surface observée) est mis en place et adapté aux acquisitions multi-angulaires de SMOS. Une comparaison entre les humidités retrouvées par ce modèle simplifié et par SMOS permettra de mieux comprendre les améliorations que SMOS est capable de fournir ainsi que celles qui pourraient lui être appliquées (chapitre 6).

Dans le dernier chapitre, une conclusion et des perspectives sur mes travaux de thèse sont présentées.

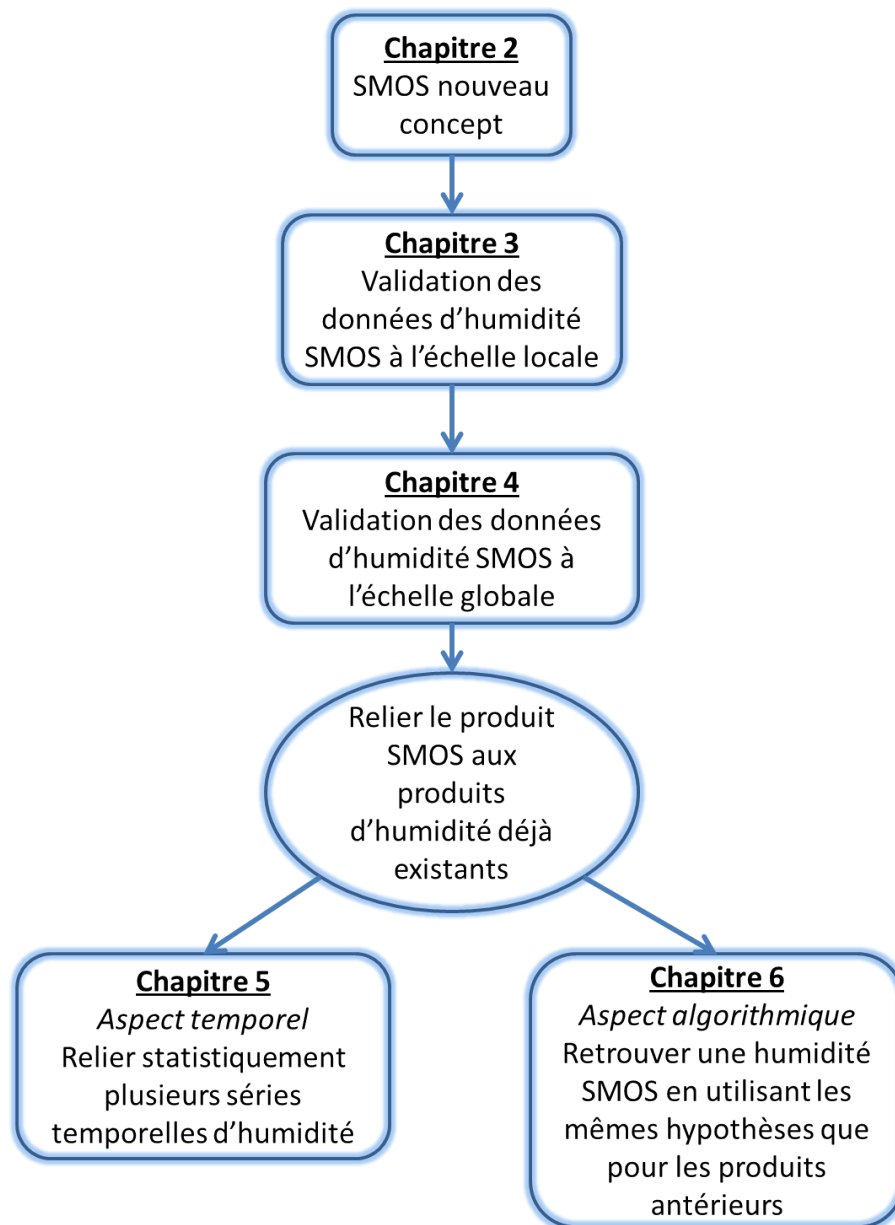


FIGURE 1.6 – Organisation de la thèse et du manuscrit.

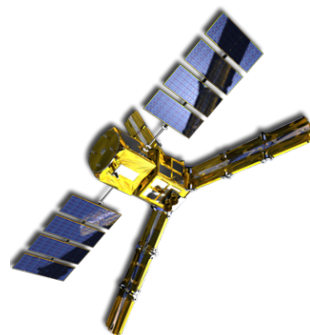
Chapitre 2

La télédétection et la mission SMOS

Dans ce chapitre est présentée tout d'abord une courte introduction à la théorie du signal micro-onde : qu'est ce qu'une onde ? comment la mesure-t-on depuis l'espace ? comment retrouver l'humidité des sols à partir de ces signaux observés ?

La mission SMOS (*Soil Moisture and Ocean Salinity*) est la première mission spécialement conçue pour observer l'humidité des sols à l'échelle globale. Le satellite a été lancé le 2 Novembre 2009 depuis la base russe de Plesetsk. Le radiomètre embarqué acquiert des informations à la fréquence la plus adéquate pour l'observation de cette variable : 1.4 GHz. De plus ces informations sont multi-angulaires, entre 0° et 55° .

Au delà de la nouveauté conceptuelle, SMOS apporte un algorithme novateur pour retrouver l'humidité et le contenu en eau de la végétation à travers la prise en compte de l'hétérogénéité de la scène observée. Les algorithmes déjà existants ne prenaient en compte qu'un seul angle d'acquisition et considéraient que l'intégralité du champ de vue du satellite était homogène à la surface. SMOS propose de découper les scènes observées en plusieurs catégories : végétation basse, forêt, zone humide, eau libre, eau salée, sol aride, glace, milieu urbain, sol enneigé et sol gelé. Cette division permet ainsi de ne retrouver l'humidité que sur les fractions d'intérêt (végétation basse et forêt) tandis que la contribution des autres fractions est estimée à travers plusieurs modèles bien spécifiques.



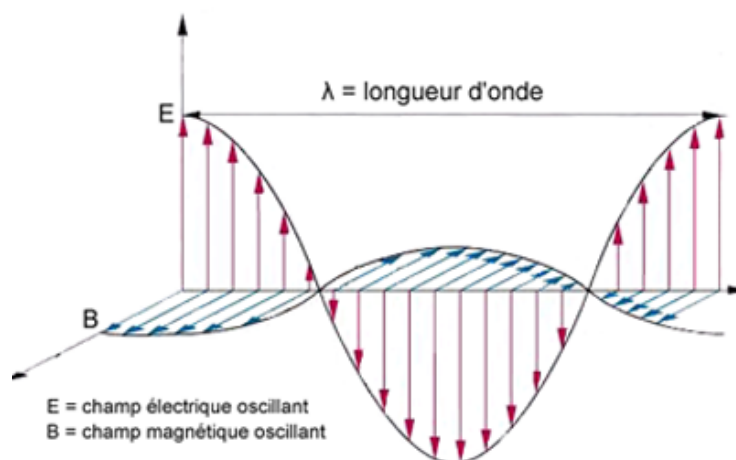


FIGURE 2.1 – Onde électromagnétique avec son champ électrique E et son champ magnétique B , source web : Bioinformatics (2002).

2.1 Théorie du signal micro-onde

2.1.1 Qu'est ce qu'une onde électromagnétique ?

Il existe deux grands types d'ondes en physique : l'onde mécanique et l'onde électromagnétique. Une onde mécanique est le déplacement d'une perturbation dans la matière (ex : le son à travers l'air, un tremblement de terre, une vague sur la mer). Ces ondes mécaniques ne peuvent se propager que dans la matière (ex : il n'y a pas de son dans l'espace). Les ondes électromagnétiques peuvent, quant à elles, se déplacer dans le vide (comme la lumière qui provient du Soleil) et dans la matière (la lumière passe à travers une vitre).

Toutes les ondes sont des déplacements d'énergie. Cependant, les ondes mécaniques et électromagnétiques diffèrent quant au type de déplacement de cette énergie. Les ondes mécaniques mettent en mouvement leur support et l'énergie est transmise par ce mouvement alors que les ondes électromagnétiques n'interagissent pas mécaniquement avec leur environnement.

Il existe en physique plusieurs types de champs. Le champ de pesanteur terrestre est un champ de force ressenti par chaque petite masse unitaire présente sur Terre. On peut définir de tels champs pour toutes les forces qui agissent à distance, et notamment les forces de type électromagnétique comme le magnétisme : on aura par exemple un champ magnétique tout autour d'un aimant, parce que chaque particule environnante dotée d'un moment magnétique ressent la force produite par cet aimant. La variation du champ électrique en un point entraîne celle du champ magnétique associé (si une boussole est placée près d'un fil parcouru par un courant électrique, l'aiguille tourne légèrement), c'est pour cela que l'on parle de champ électromagnétique.

Une onde électromagnétique est une vibration de ce champ électromagnétique. La modification de la valeur du champ électrique E entraîne une modification de celle du champ magnétique B et vice-versa (fig. 2.1). Finalement, les champs électrique et magnétique changent de valeur alternativement et l'énergie utilisée pour créer la première variation du champ électromagnétique est transportée à chaque variation successive. Une onde électromagnétique est donc la propagation de variations (ou signal) grâce à un champ électrique et un champ magnétique qui vibrent ensemble.

Une onde est caractérisée par deux paramètres indépendants : sa longueur d'onde (ou sa fréquence) et sa vitesse de propagation. La longueur d'onde λ est la distance (mesurée en

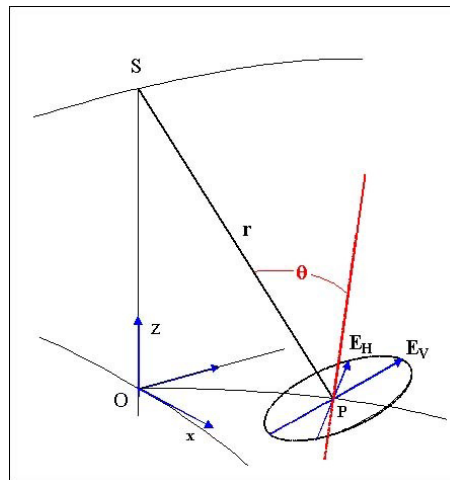


FIGURE 2.2 – Définition des polarisations verticales et horizontales : V dans le plan perpendiculaire au plan d'incidence et H dans le plan parallèle à la surface de la Terre (ATBD, 2010).

mètre) entre deux oscillations successives, ou entre deux points de l'onde qui sont dans le même état de mouvement. La fréquence (mesurée en hertz - Hz) est le nombre de perturbations ou d'oscillations qui passent en un endroit donné en une seconde. L'inverse de la fréquence, appelée période, est le temps (mesuré en secondes) nécessaire pour que l'onde parcoure un cycle. Dans le cas d'une onde électromagnétique, la vitesse de propagation dans le vide est égale à la vitesse de la lumière dans le vide : $3 \cdot 10^8$ m/s.

Une onde électromagnétique est donc un transfert d'énergie. Tout corps ayant une température supérieure à 0 kelvin (-273.15°C) émet un rayonnement électromagnétique appelé également rayonnement thermique ou rayonnement du corps gris. Le corps noir est un objet idéal qui absorberait toute l'énergie électromagnétique qu'il recevrait, sans en réfléchir ni en transmettre. La lumière étant un rayonnement électromagnétique, elle est absorbée totalement et l'objet devrait donc apparaître noir, d'où son nom.

En télédétection, l'onde électromagnétique se décompose en deux vecteurs orthogonaux entre eux et orthogonaux à la direction de propagation qui sont appelés polarisation verticale et polarisation horizontale (fig. 2.1). Par convention, la polarisation verticale est celle contenue dans le plan perpendiculaire au plan d'incidence ou d'observation et qui contient la direction de propagation, et la polarisation horizontale est contenue dans le plan parallèle à la surface de la Terre (fig. 2.2).

2.1.2 Comment cela se mesure-t-il ?

Il existe deux types d'instruments micro-ondes : passifs et actifs. Les capteurs passifs, ou radiomètres, mesurent l'énergie micro-onde qui est émise par émission thermique ou réfléchi par la surface de la Terre et l'atmosphère. Une fine sélection de la fréquence d'observation permet de mesurer les contributions de l'atmosphère, du sol, de la glace ou de la pluie. Les radiomètres utilisés pour mesurer les émissions atmosphériques sont plus particulièrement appelés sondeurs micro-ondes (quand ils donnent un profil de l'atmosphère) ou imageurs.

Les capteurs actifs, comme les systèmes radar, génèrent leur propre illumination en transmettant des impulsions d'ondes micro-ondes puis en utilisant un système récepteur pour mesurer le signal réfléchi (ou diffusé) par la zone d'intérêt visée. Si l'objectif de la mission est l'étude de la surface terrestre, ces instruments utilisent essentiellement des fréquences inférieures à 10 GHz où l'atmosphère devient quasiment invisible (comme déjà montré sur la figure 1.5).

Une partie du domaine micro-onde permet d'observer la Terre à travers les nuages (idéal pour observer des régions perpétuellement couvertes par les nuages, comme par exemple, les régions tropicales humides). Elles permettent également l'observation de jour comme de nuit, ce qui offre l'opportunité d'obtenir une couverture continue par les instruments micro-ondes embarqués sur satellites.

2.1.2.1 Les imageurs micro-ondes passifs

L'observation de la Terre par micro-onde passive offre une représentation unique de sa surface et de l'atmosphère qui l'entoure. Les mesures sont en effet sensibles aux variations du contenu en eau de l'atmosphère, de la température de la surface terrestre, de sa rugosité, salinité et teneur en eau. Les trois derniers paramètres sont d'autant plus intéressants qu'ils ne peuvent être observés dans aucune autre gamme de fréquence. Les capteurs actifs peuvent également observer ces variables mais ils sont en général plus volumineux et nécessitent plus de puissance pour fonctionner et transmettre un plus gros volume de données (Woodhouse, 2006). Les satellites avec des imageurs micro-ondes passifs (ou radiomètres imageurs) représentent donc un bon compromis concernant la masse et la puissance, qui représentent des points clés lors du financement d'un lancement de satellite. L'impact minime de la couverture nuageuse et des aérosols représente un grand avantage pour une partie de la région micro-onde.

Le défi dans l'utilisation des micro-ondes passives se situe au niveau de la résolution d'observation puisque les longueurs d'onde sont grandes par rapport à la taille des antennes. Pour des systèmes embarqués à bord de satellite, la résolution est de plusieurs dizaines de kilomètres, ce qui limite l'utilisation de ces observations aux échelles régionales et globales.

2.1.2.2 Les capteurs micro-ondes actifs

Comme les imageurs micro-ondes passifs, les capteurs micro-ondes actifs sont sensibles aux variations de rugosité, salinité et teneur en eau. Cependant, la technologie utilisée n'est pas la même puisque l'instrument génère lui même sa propre source d'illumination qu'il va envoyer en direction de la Terre pour ensuite mesurer l'écho de cette onde qui a été réfléchi et dissipé par la surface. L'étude de cet écho représente une information supplémentaire puisque le temps de réponse est enregistré, ce qui permet d'estimer la distance qui sépare l'instrument de sa cible (méthode utilisée pour réaliser des modèles topographiques de surface). De plus, cet écho revient avec une variation d'intensité et de polarisation par rapport au signal préalablement envoyé. Ces variations correspondent à des variations des paramètres physiques de la surface observée. Les instruments se focalisant sur l'étude des délais des échos sont appelés altimètres alors que ceux qui étudient les variations en intensité et polarisation sont des diffusiomètres.

On peut également noter l'existence des radars à synthèse d'ouverture (SAR) qui permettent d'améliorer grandement la résolution sans augmenter la taille des antennes. En effet, le faisceau est conçu pour être plus large et pour que les empreintes successives de l'instrument se chevauchent. Chaque point est observé plusieurs fois et en combinant les différentes empreintes, il est possible de *synthétiser* une antenne plus grande (Woodhouse, 2006).

Mis à part les systèmes SAR, les systèmes actifs micro-ondes ont les mêmes limitations que les capteurs passifs en terme de résolution, les grandes longueurs d'onde sont automatiquement associées à de grandes antennes avec des résolutions de l'ordre de quelques kilomètres. Les données sont alors plus particulièrement utilisées pour les études à l'échelle régionale ou globale.

2.1.3 Comment retrouver l'humidité du sol à partir d'un signal ?

Les rayonnements électromagnétiques observés et mesurés par les instruments passifs sont en fait reliés à des notions de brillance et de température. Ce sont les variations de ces températures qui permettent au final de retrouver des valeurs de paramètres physiques tels que l'humidité du sol.

2.1.3.1 Notion de température de brillance pour les micro-ondes passives

Pour quantifier le rayonnement électromagnétique qu'émet un objet ou une scène, on définit la notion de brillance ou de luminance qui représente la quantité d'énergie par unité de surface émise à une fréquence donnée (Ulaby et al., 1986). Pour cela la loi de Planck définit mathématiquement la brillance de la manière suivante :

$$I_\nu = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \quad (2.1)$$

avec :

- I la brillance ou luminance ($\text{W} \cdot \text{s} / (\text{m}^2 \cdot \text{sr})$)
- ν la fréquence (Hz)
- h la constante de Planck ($6.63 \cdot 10^{-34} \text{W} \cdot \text{s}^2$)
- c la vitesse de la lumière ($3 \cdot 10^8 \text{m/s}$)
- k la constante de Boltzmann ($1.38 \cdot 10^{-23} \text{W} \cdot \text{s/K}$)
- T la température du corps noir (K)

Cette loi peut être simplifiée avec l'hypothèse qu'aux faibles fréquences et avec des hautes températures, on a $h\nu \ll kT$ (loi de Rayleigh-Jeans) qui définit la brillance I proportionnelle à la température T :

$$I_\nu = \frac{2\nu^2 k}{c^2} T \quad (2.2)$$

Le corps noir est un concept idéal où l'objet absorbe et émet toute l'énergie électromagnétique reçue. La brillance d'un corps noir I_{cn} pour une largeur de bande $\Delta\nu$ à une température T est définie par :

$$I_{cn} = I_\nu \cdot \Delta\nu = \frac{2\nu^2 k}{c^2} T \Delta\nu \quad (2.3)$$

Il n'existe pas de tel objet en réalité. On parle alors de corps gris qui présente des pertes par rapport au corps noir (leur capacité à absorber et à émettre est inférieure à celle d'un corps noir). On définit la brillance d'un corps gris I_{cg} à une température T pour une largeur de bande $\Delta\nu$ de la même manière que pour le corps noir :

$$I_{cg} = \frac{2\nu^2 k}{c^2} T_B \Delta\nu \quad (2.4)$$

avec T_B la température de brillance du corps gris qui est toujours inférieure à la température T (pour un corps noir $T_B = T$). T_B est la température qu'aurait un corps noir s'il rayonnait la même énergie que le corps gris (sans les pertes). On peut définir alors l'émissivité e qui est le rapport entre ces deux températures et qui est toujours inférieur ou égal à 1 :

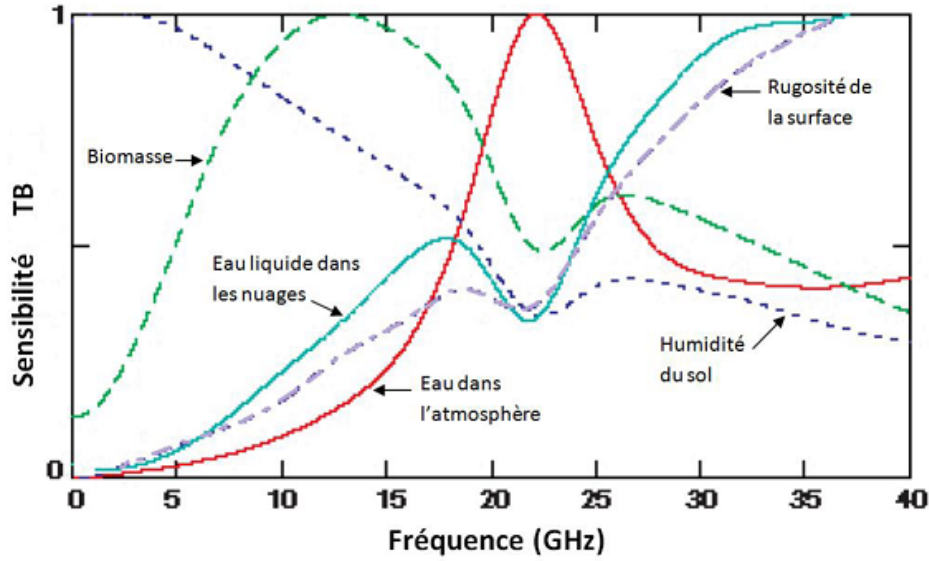


FIGURE 2.3 – Sensibilité de la température de brillance à la biomasse, à l’eau liquide dans les nuages, à l’eau dans l’atmosphère, à l’humidité du sol et à la rugosité de surface en fonction de la fréquence (Kerr, 1996).

$$e = \frac{I_{cg}}{I_{cn}} = \frac{T_B}{T} \quad (2.5)$$

En réalité, la température de brillance et l’émissivité dépendent de la direction d’émission ou d’observation, mais aussi de plusieurs paramètres physiques.

2.1.3.2 Retrouver l’humidité à partir des températures de brillance

Les températures de brillance mesurées par les radiomètres sont en réalité le résultat des conditions de la surface observée (en mettant de côté l’aspect directionnel de la température de brillance). Le rayonnement électromagnétique ne sera en effet pas le même au dessus d’un désert, d’une forêt ou d’une plaine agricole. C’est cette différence qui permet de retrouver certains paramètres de la surface. De plus, la fréquence d’observation est très importante car la sensibilité du rayonnement ou de la température de brillance varie selon les paramètres et la fréquence.

La figure 2.3 montre la sensibilité de la température de brillance en fonction de la fréquence pour chacun des paramètres biophysiques suivants : la biomasse (végétation), l’eau liquide contenue dans les nuages, l’eau dans l’atmosphère, l’humidité du sol et la rugosité de la surface. On peut ainsi identifier les fréquences les plus adaptées pour l’observation de chaque variable : en dessous de 3 GHz pour l’humidité du sol, 13 GHz pour la biomasse, 22 GHz pour le contenu en eau de l’atmosphère, et au delà de 35 GHz pour le contenu en eau liquide des nuages et pour la rugosité.

Cependant, même si une fréquence est la plus adaptée pour l’observation d’un paramètre, cela ne veut pas dire que l’évolution de la température de brillance est entièrement due à l’évolution de ce paramètre. A 22 GHz, la température de brillance est très sensible à la vapeur d’eau dans l’atmosphère mais elle reste également sensible à tous les autres paramètres. Néanmoins, si on connaît ou si on sait estimer tous les paramètres sauf un, il suffit alors de sélectionner la fréquence adéquate.

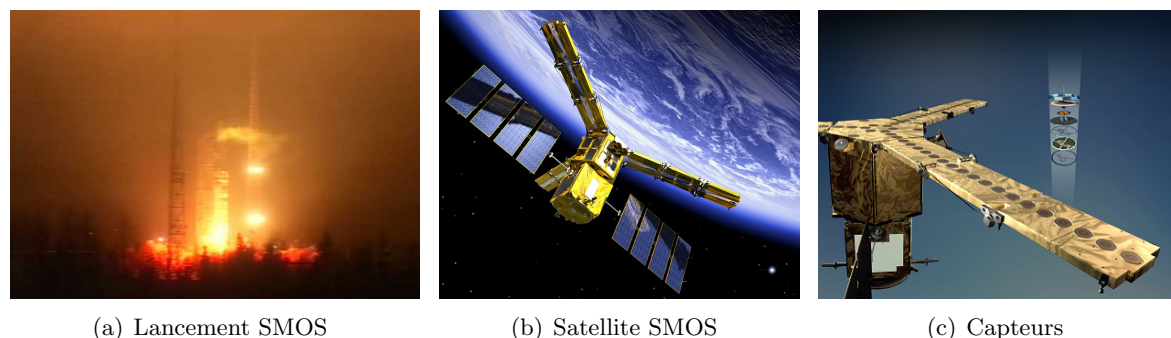


FIGURE 2.4 – Le satellite SMOS lors du lancement à Plesetsk le 2 Novembre 2009 (a) avec une représentation du satellite en orbite (b) ainsi qu’une vue détaillée d’un receveur (c) qui permet de capter les ondes électromagnétiques émises par la Terre. Images ESA.

A des fréquences inférieures à 3 GHz, la température de brillance est majoritairement influencée par la salinité des eaux et à l’humidité des sols mais elle reste cependant sensible à la végétation. Si on sait estimer l’effet de la végétation sur le rayonnement émis, alors on pourra retrouver l’humidité du sol. De plus la bande L (1-2 GHz) est une bande protégée et réservée pour une utilisation strictement scientifique. Cependant, cette protection n’est pas toujours respectée dans certaines parties du monde, notamment en Asie.

2.2 La mission SMOS

La mission SMOS (*Soil Moisture and Ocean Salinity*, Kerr et al. (2010)) fait partie des Missions d’Opportunités d’Exploration de la Terre menées par l’Agence Spatiale Européenne (ESA) et a été lancée le 2 Novembre 2009. Le premier objectif de la mission SMOS est l’observation globale de l’humidité des sols et de la salinité des océans. Le lancement a été effectué depuis la base de Plesetsk dans le Nord de la Russie (figure 2.4(a)).

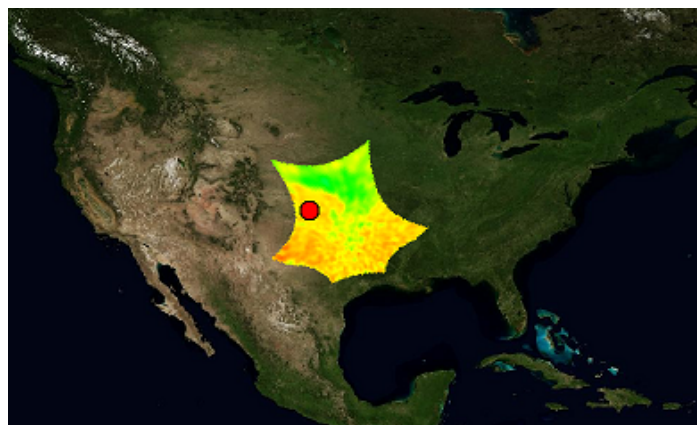
Les objectifs scientifiques de cette mission sont (1) de mesurer l’humidité des sols avec une précision de $0.04 \text{ m}^3/\text{m}^3$ (correspondant aux premiers centimètres sous la surface) et une résolution spatiale de moins de 50 km avec une couverture globale tous les trois jours, (2) de fournir des cartes de salinité globales de l’eau en surface des océans à une précision de 0.1 PSU (*Practical Salinity Unit*) avec une résolution spatiale de 200 km tous les dix jours ($1 \text{ psu} = 1 \text{ g de sel par kg d’eau de mer}$).

Le satellite SMOS suit une orbite circulaire héliosynchrone (le satellite repasse au dessus d’un point donné de la Terre toujours à la même heure) à 755 km d’altitude. Dans sa trajectoire ascendante (du Sud vers le Nord), le satellite passe l’équateur à 6h et dans sa trajectoire descendante (du Nord vers le Sud) à 18h. Cette mission a été programmée pour un minimum de trois ans avec une possible extension de deux ans.

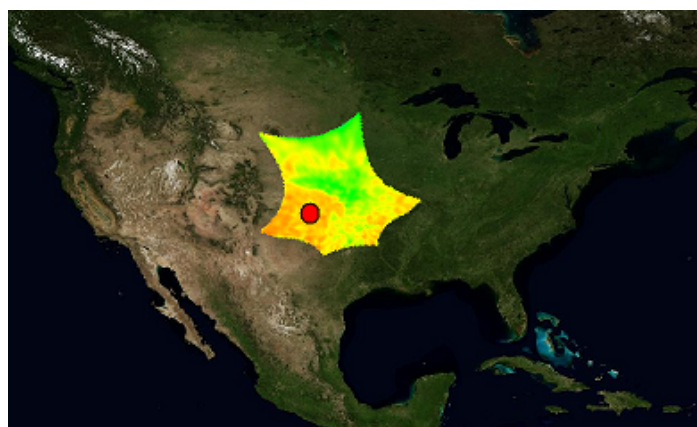
2.2.1 L’instrument

L’instrument SMOS est composé d’un radiomètre pour micro-ondes passives (1.41 GHz) constitué d’une structure centrale et de trois bras (figure 2.4(b)). Il est incliné à 32.5° par rapport au nadir (i.e. direction située directement en dessous du satellite) pour garantir une gamme d’angle d’incidence de 0° à 55° sur Terre.

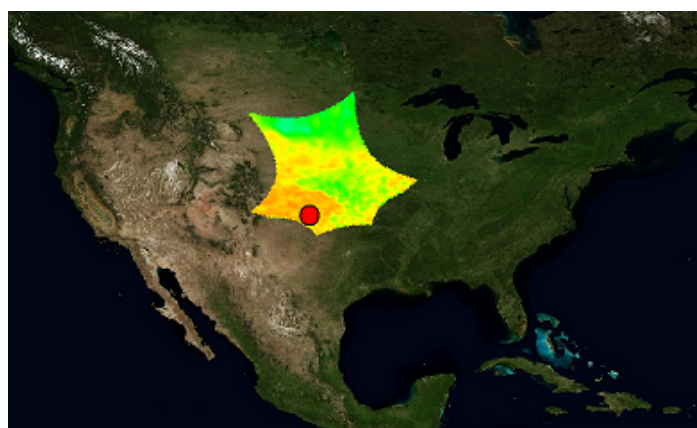
Le principe de cet instrument repose sur un ensemble de 69 récepteurs (ou antennes élémentaires) qui sont répartis régulièrement sur les trois bras du satellite. La corrélation entre le signal de sortie de tous les couples de récepteurs donne un estimateur de la fonction de visibilité. Ces



(a) Snapshot 1



(b) Snapshot 2



(c) Snapshot 3

FIGURE 2.5 – Données SMOS de niveau 1 organisées par images acquises par le radiomètre appelées snapshots. Un point (indiqué en rouge) est donc vu à plusieurs reprises par les capteurs à différents angles d'incidence.

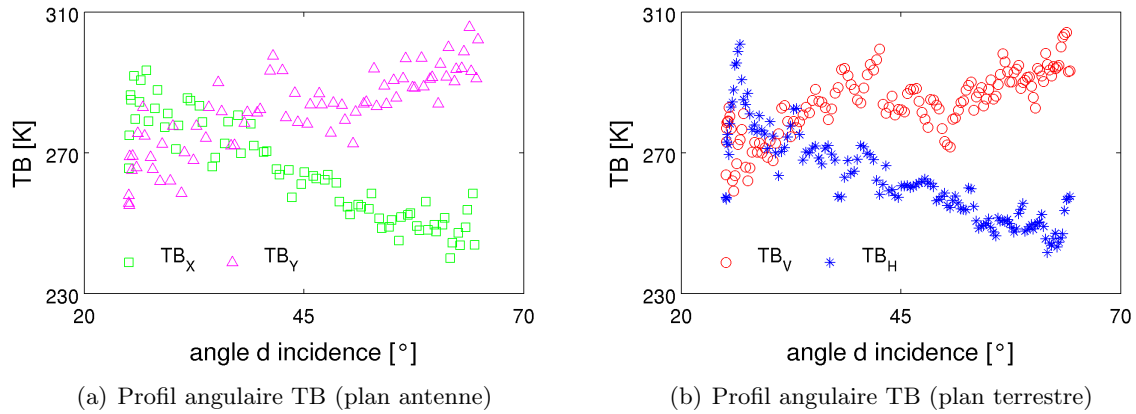


FIGURE 2.6 – Données SMOS de niveau 1c : températures de brillance TB dans le référentiel de l’antenne (a - polarisations X et Y) auxquelles il faut appliquer une rotation pour les obtenir dans le référentiel terrestre (b - polarisations Horizontale et Verticale).

fonctions de visibilité sont ensuite transmises au sol où une opération (pseudo transformée de Fourier inverse) est effectuée afin de reconstruire l’image des températures de brillance correspondant à la scène observée (Anterrieu & Khazaal, 2008). Cette méthode de corrélation entre les signaux est appelée interférométrie.

L’interférométrie permet également d’obtenir des observations avec une meilleure résolution qu’avec des antennes circulaires. Pour une résolution équivalente, une antenne circulaire de 9 m aurait été nécessaire, ce qui ne convient pas en terme de poids et de taille pour un lancement satellite. Chaque bras du satellite SMOS mesure 4 m de longueur pour 25 cm de large. Pour le lancement, ces bras ont été repliés puis déployés une fois la mise en orbite effectuée.

2.2.2 Les données SMOS

SMOS génère plusieurs niveaux de données, correspondant à plusieurs niveaux de traitement. Le niveau 0 (L0) représente les données brutes provenant de l’instrument. Le niveau 1 est découpé en trois sous-niveaux de traitement. Le niveau 1a (L1a) contient les visibilité (corrélations entre les receveurs), le niveau 1b (L1b) est le résultat de la reconstruction de l’image à partir des visibilité : les températures de brillance dans le plan de l’antenne ; chaque *image* prise par l’instrument est appelée *snapshot*. Les données L1b ont été organisées par snapshot (figure 2.5). Enfin le niveau 1c (L1c) est une réorganisation géographique du niveau 1b qui permet d’obtenir pour chaque point toutes les températures de brillance acquises à différents angles d’incidence (appelé profil angulaire). Ces températures de brillance sont stockées par demi-orbite et sont encore dans le référentiel de l’antenne (figure 2.6(a)). Une opération est donc nécessaire pour projeter ces données du plan de l’antenne au plan terrestre (correction des effets dus au passage des ondes électromagnétiques dans la ionosphère et à l’attitude du satellite, figure 2.6(b)).

Le niveau 2 (L2) représente le passage à l’humidité du sol ou à la salinité des océans. Les algorithmes sont appliqués aux données L1c (températures de brillance aux différents angles d’incidence) et utilisent les différences ou signatures angulaires pour retrouver ces variables. Des données auxiliaires sont également utilisées pour connaître l’état et l’occupation des sols, par exemple une zone enneigée ou une zone à fort relief. Ces données sont stockées par demi-orbite. L’historique des versions du produit d’humidité de niveau 2 est présenté en annexe G (en anglais). Le produit d’humidité de niveau 2 est une humidité instantanée et pour un point donné,

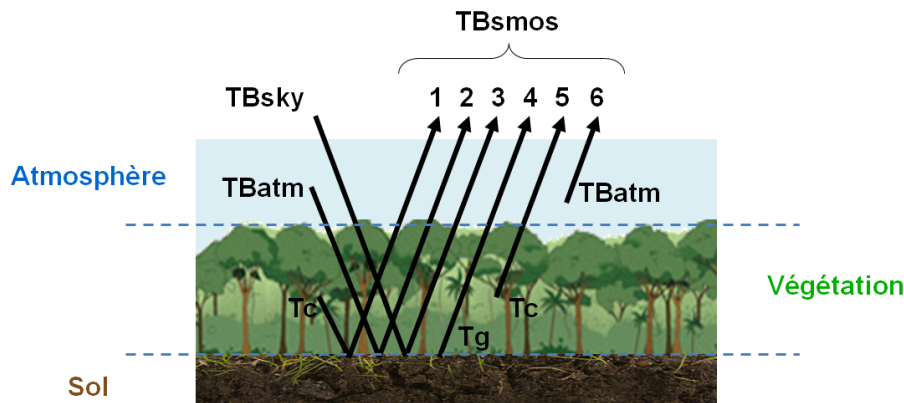


FIGURE 2.7 – Modèle de transfert radiatif avec les différentes contributions de l’atmosphère, de la végétation et du sol. Le radiomètre SMOS observe la somme de ces contributions et afin de retrouver l’humidité du sol, il faut pouvoir estimer certaines de ces contributions.

une humidité devrait être retrouvée au moins tous les trois jours si les conditions algorithmiques et physiques sont réunies.

Le niveau 3 (L3) est une globalisation du niveau 2. Des cartes globales sont ainsi générées tous les 3 jours. Le niveau 4 (L4) est une combinaison des données SMOS avec des données extérieures, permettant d’aboutir à une nouvelle connaissance : l’humidité en zone racinaire (1 mètre de profondeur), prévention de risques naturels (feu, inondation, sécheresse) ou encore l’épaisseur de la glace.

2.2.3 L’algorithme d’inversion des données SMOS

L’innovation de la mission SMOS se situe en premier lieu dans l’observation de la Terre d’une façon multi-angulaire et dans une gamme de fréquence encore jamais explorée à l’échelle globale. Le radiomètre embarqué sur SMOS retranscrit des températures de brillance acquises à divers angles d’incidence. Ces profils angulaires de températures sont ensuite comparés au modèle L-MEB (*L-band Microwave Emission of the Biosphere*, Wigneron et al. (2007)) afin de retrouver l’humidité sur chaque point de la grille (Kerr et al., 2012, ATBD, 2010). On appelle ce procédé inversion de l’humidité.

Le modèle L-MEB est basé sur un modèle de transfert radiatif qui comporte trois sources d’émission de signal : l’atmosphère, la végétation et le sol (fig. 2.7). Le signal acquis par SMOS provient ainsi (6) de l’atmosphère, (5) de la végétation, (4) du sol, (3) du rayonnement galactique traversant l’atmosphère et la végétation réfléchi par le sol, (2) de l’atmosphère traversant la végétation puis réfléchi par le sol et (1) de la végétation réfléchi par le sol (fig. 2.7). Pour retrouver l’humidité du sol et parfois l’opacité de la végétation, il faut être en mesure de bien estimer les autres paramètres atmosphériques, l’occupation du sol ou encore la texture du sol qui donnera les capacités d’émission et de réflexion du sol. Avec ces éléments auxiliaires, le modèle L-MEB est capable de simuler des températures de brillance pour tous les angles d’incidence. L’humidité du sol et l’épaisseur optique de la végétation sont retrouvées de telle façon que la différence entre les températures de brillance acquises par SMOS et simulées par L-MEB soit la plus faible possible.

Une autre innovation qu’apporte le produit SMOS se situe dans l’algorithme lui même. Les algorithmes déjà existants posent une hypothèse très forte concernant le champ de vue de l’instrument : toute scène observée est supposée homogène (fig. 2.8(b)). Cela veut dire que

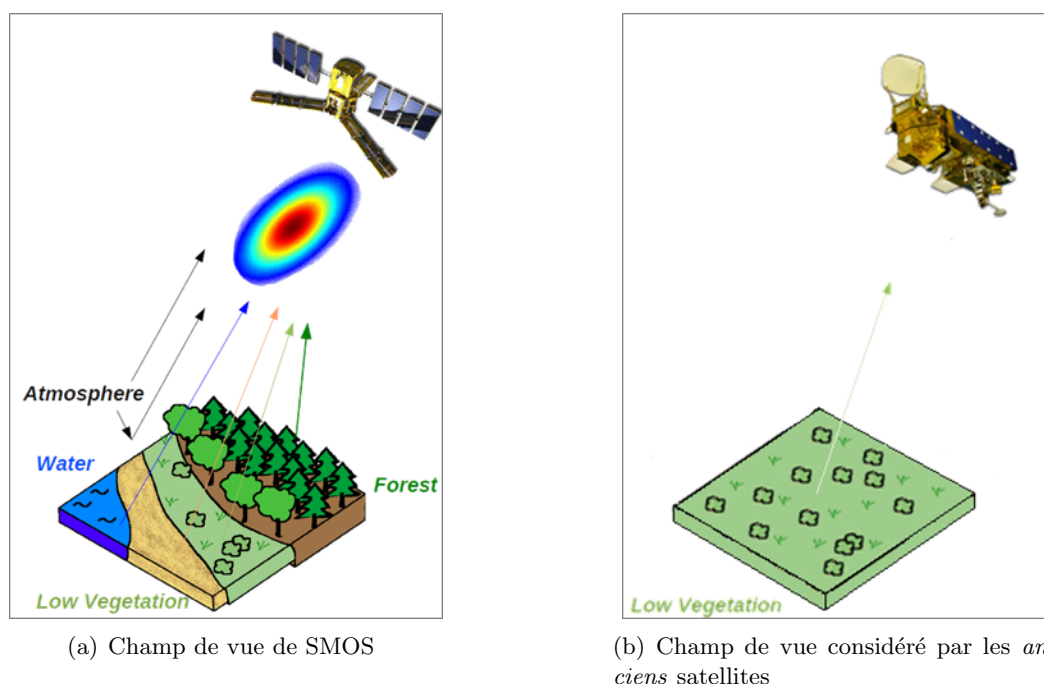


FIGURE 2.8 – Différences entre les considérations de SMOS et des satellites comme ASMR-E concernant leurs champs de vue : SMOS prend en compte l'hétérogénéité de la surface en fractionnant le champ de vue et y applique une pondération correspondant au diagramme d'antenne alors que les satellites comme AMSR-E considèrent une surface homogène.

même si cette zone est constituée de forêts, de plaines agricoles et de villes, ces éléments ne sont pas différenciés et ne sont pas traités différemment. Or avec une résolution instrumentale de 50 km, c'est une hypothèse très forte qui peut conduire à des approximations très importantes, notamment dans les zones de transition (côtes ou forêts).

Dans l'algorithme de SMOS, chaque scène observée est analysée afin d'identifier les différents types d'occupation des sols (fig. 2.8(a)). Pour chaque catégorie, un ensemble de paramètres est défini et il est ainsi possible d'identifier les contributions correspondantes à chacune de ces catégories et de ne retrouver l'humidité que sur la partie où il est intéressant d'avoir cette information (l'humidité des sols n'a pas de sens dans une zone urbaine ou sur des marécages).

Lorsque le radiomètre *regarde* dans une direction, ou vise un point, il n'est pas seulement influencé par ce point, il est également influencé par ce qui entoure ce point. Pour cela, une surface carrée de $123 \times 123 \text{ km}^2$ (appelée surface de travail qui correspond à la surface maximale observée par l'instrument) est définie autour de chaque point où l'on souhaite retrouver l'humidité. La base de données d'occupation des sols ECOCLIMAP (Masson et al., 2003) est ensuite utilisée afin d'identifier la composition de cette surface de travail. ECOCLIMAP a une résolution de $30''$ d'arc (un peu moins d'1 km à l'équateur) et est composée de 256 types d'occupation. Ces 256 classes ont été regroupées en 8 classes plus générales sur une grille moins fine (pour des raisons de temps de calcul, environ 4km de résolution) : végétation basse, forêt, zone humide, eau libre, eau salée, sol aride, glace et milieu urbain. Ainsi, une carte globale de l'occupation des sols est disponible comprenant 8 classes agglomérées. Un exemple du bassin de la Garonne est présenté en figure 2.9. Il est important de noter que cette base de données ECOCLIMAP est fixe et n'évolue pas au cours des saisons. Elle n'est donc peut être pas exacte à un temps t précis.

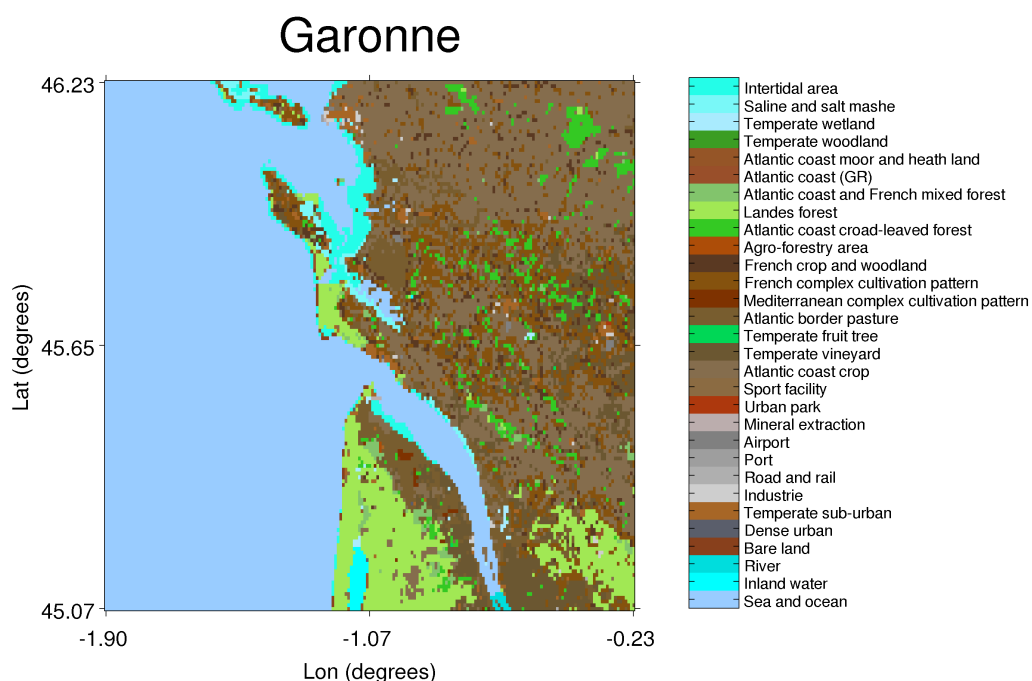
Afin de mieux représenter ce qui est réellement vu par le radiomètre, une fonction de poids

est appliquée à la surface de travail. Cette fonction de poids dépend de l'angle d'incidence et de la position du point considéré dans le champ de vue de l'instrument. Ainsi il est possible de déterminer les fractions que représente chaque catégorie présente dans le champ de vue de l'instrument et d'en déduire leur contribution au signal. L'humidité n'est retrouvée que sur les zones à végétation basse et les forêts. Les contributions au signal des autres catégories (zone humide, eau libre, eau salée, sol aride, glace et milieu urbain) sont estimées grâce à des modèles et des paramètres spécifiques.

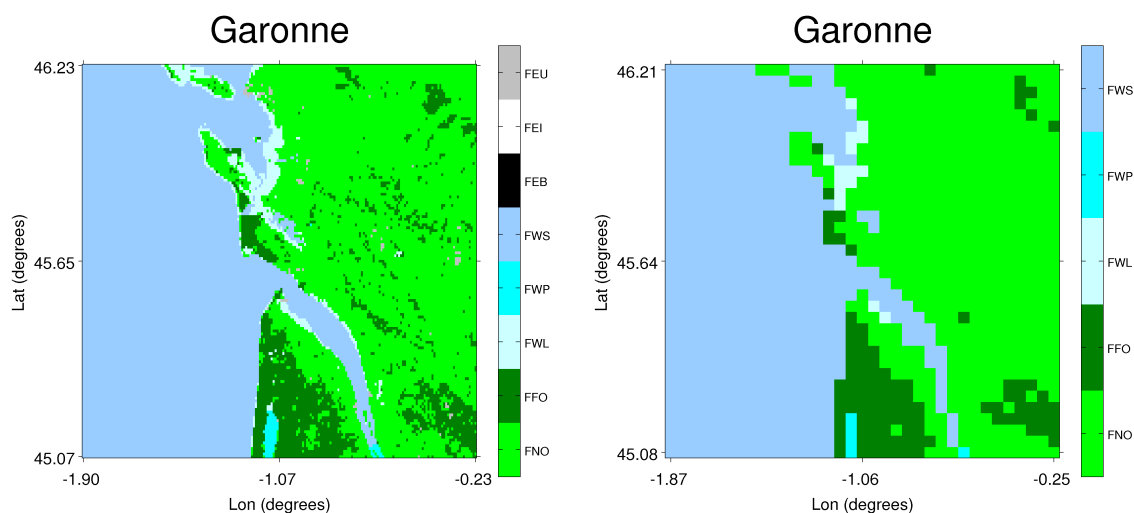
Prenons un exemple : une scène est constituée de 75% de végétation basse, 20% de zone humide et 5% de villes, et pour retrouver l'humidité sur la zone à faible végétation, il faut d'abord estimer la contribution au signal des deux autres catégories (en utilisant leurs modèles respectifs) afin de retrouver la part de la température de brillance acquise qui correspond à la zone à faible végétation. Une fois les contributions identifiées, il est possible de retrouver l'humidité du sol à faible végétation. De l'eau libre, comme un fleuve, aura tendance à diminuer la température de brillance. Ce procédé permet ainsi de prendre en compte l'hétérogénéité de l'occupation des sols. Les différents modèles des différentes classes d'occupation des sols sont détaillés dans l'ATBD (*Algorithm Theoretical Basis Document*) de niveau 2 de SMOS (ATBD, 2010).

Il existe cependant des limites à cette inversion de l'humidité du sol, notamment sur les zones à forte végétation. Par exemple les forêts équatoriales humides sont très complexes puisqu'elles représentent une couche épaisse de végétation et le signal observé par l'instrument provient essentiellement de cette couche de végétation et le sol est quasiment invisible. En revanche, il est parfois possible de retrouver l'humidité sur des forêts moins denses telles que les forêts boréales à pins ou arbres à aiguilles. La modélisation de la litière présente sous-bois est également importante dans la modélisation de la contribution de la forêt comme le souligne les travaux de Bircher et al. (2012).

Le chapitre 6 détaille comment l'humidité est retrouvée à partir des profils angulaires des températures de brillance et comment l'humidité et la végétation influent sur les observations.



(a) Zone de travail représentée par les 256 classes d'ECOCLIMAP



(b) Zone de travail représentée avec les 8 classes génériques

(c) Zone de travail à basse résolution

FIGURE 2.9 – Exemple d'une zone de travail représentée (a) à travers 30 classes ECOCLIMAP (il en existe 256 en tout) à une résolution de 30" d'arc; (b) les classes ECOCLIMAP ont été regroupées en 8 classes génériques : basse végétation (FNO), forêt (FFO), zone humide (FWL), eau pure (FWP), eau salée (FWS), zone aride (FEB), glace (FEI) et zone urbaine (FEU); (c) les 8 classes génériques à une résolution de 4km.

Chapitre 3

Évaluation du produit SMOS à l'échelle locale

Dans le cadre de la validation des données d'humidité de la mission SMOS (*Soil Moisture and Ocean Salinity*), une comparaison entre trois jeux de données satellite, modèle et terrain est menée sur quatre bassins versants test aux Etats-Unis. Quatre produits d'humidité sont comparés en 2010 avec des mesures d'humidité effectuées par des réseaux de stations météorologiques qui ont déjà été utilisés pour la validation et la calibration des données du satellite AMSR-E (*Advanced Microwave Scanning Radiometer*).

Les résultats de cette comparaison montrent que les humidités retrouvées par SMOS sont plus proches des mesures terrain avec une faible erreur totale (RMSE) de $0.052 \text{ m}^3/\text{m}^3$ pour le matin et de $0.073 \text{ m}^3/\text{m}^3$ pour l'après-midi, ce qui représente une amélioration d'un facteur 2 à 3 comparé aux autres produits d'humidité existants. Le produit ECMWF (*European Center for Medium range Weather Forecasting*) est très bien corrélé avec les mesures terrain mais a un biais constant entre 0.1 et $0.2 \text{ m}^3/\text{m}^3$ sur les quatre sites test. Le produit VUA (*Vrije University of Amsterdam*) donne des résultats satisfaisants mais présente un biais variant avec les saisons qui est corrélé avec le développement de la végétation et la température du sol. L'indice d'humidité ASCAT (*Advanced Scatterometer*) est très bruité et instable sur les quatre bassins versants.



3.1 Introduction

3.1.1 Objectifs et défis

Lors de la mise à disposition d'un nouveau jeu de données, une phase de validation doit tout d'abord être respectée afin d'évaluer au mieux ce nouveau produit : identifier ses forces, ses faiblesses et les comparer aux produits déjà existants. Dans un premier temps, on utilise des données provenant de stations météorologiques équipées de sondes qui mesurent l'humidité contenue dans le sol et qui sont généralement disposées dans les 5 à 7 premiers centimètres sous la surface du sol. Ainsi les humidités déduites des observations du satellite SMOS peuvent être comparées aux mesures de ces stations.

Cependant, lors de cette comparaison, plusieurs défis sont à relever :

- i Pour les radiomètres à micro-ondes passives, la taille de la zone observée ou couverte par le satellite (*footprint* en anglais ou empreinte au sol en français) n'est pas précisément connue et n'est que rarement prise en compte. Les diagrammes d'antenne, qui ont été définis lors de la fabrication de l'instrument, permettent d'estimer d'où provient la majorité du signal (au moins 50% en terme de puissance de signal reçu).
- ii La plupart des radiomètres, et d'autant plus SMOS qui observe avec plusieurs angles d'incidence avec 69 antennes élémentaires, observe une surface qui n'est pas un cercle, mais plutôt une ellipse (intersection du cône d'observation et de la Terre considérée sphérique) et lorsque le satellite observe cette même zone quelques jours plus tard, il ne l'observera pas de la même manière ou à travers une ellipse différente. Par conséquent la forme de l'empreinte évolue lors du passage du satellite. De ce fait, un réseau de stations mesurant l'humidité du sol ne peut être conçu idéalement pour tous les passages de chaque satellite.
- iii L'épaisseur de sol qui contribue au signal varie avec la fréquence d'observation (plus la fréquence est élevée, moins l'épaisseur est importante) et avec la valeur d'humidité elle-même (plus le sol est sec, plus l'épaisseur est importante). Ainsi, à 1.4 GHz quand le sol est sec, les observations satellites peuvent être représentatives de plusieurs dizaines de centimètres sous la surface du sol (cas d'un sol très sableux) alors qu'à 6 GHz quand le sol est humide, elles ne représentent finalement que les tout premiers centimètres voire moins.
- iv Puisque les empreintes au sol en micro-onde passive mesurent des dizaines de kilomètres (jusqu'à plus de 100 km si l'angle d'observation est élevé), les surfaces observées peuvent être très hétérogènes en terme de texture et d'occupation des sols (forêts, cultures, zones urbaines,...), une valeur unique d'humidité ne peut pas être représentative d'une surface entière. Cela peut conduire à de grandes différences lors des comparaisons.
- v La variabilité de l'humidité est naturellement très forte. Il n'est pas rare d'observer des différences de l'ordre de 50% dans un rayon d'un mètre ou dans une même journée et il est donc réellement difficile de représenter une surface de plusieurs kilomètres.

L'objectif de ce chapitre est de valider les données d'humidité SMOS sur un réseau de stations météorologiques qui représente au mieux le champ de vue du satellite et qui soit représentatif de plusieurs écosystèmes afin de tester les performances de SMOS sous différentes conditions climatiques. Dans un premier temps, différents produits d'humidité retrouvés à partir d'observations satellite et de modèle auxquels les produits SMOS sont comparés sont présentés, suivi d'une description des quatre bassins versants test aux États-Unis sur lesquels se déroule cette étude. La section 3.3 présente les résultats de cette comparaison entre les différents produits d'humidité et les mesures relevées par les stations météorologiques. La dernière section présente les conclusions et les perspectives de la phase d'évaluation et de validation des données SMOS à l'échelle locale.

3.1.2 Données satellites

3.1.2.1 L'instrument AMSR-E

L'*Advanced Microwave Scanning Radiometer – Earth observing system* (AMSR-E, Kawanishi et al. (2003)) fut lancé en Juin 2002 à bord du satellite Aqua et cessa de fonctionner en Octobre 2011. Ce radiomètre observait la Terre avec un angle d'incidence unique d'environ 55° à 6 fréquences différentes : 6.9, 10.7, 18.7, 23.8, 36.5 et 89.0 GHz (polarisations horizontale et verticale). Les heures de passage à l'équateur étaient de 01:30 (heure solaire locale, HSL) pour l'orbite descendante et de 13:30 (HSL) pour l'orbite ascendante.

De nombreux produits d'humidité du sol utilisant les observations d'AMSR-E sont disponibles. Le produit officiel est le produit NSIDC (*National Snow and Ice Data Center*, Njoku et al. (2003)). Plusieurs études ont déjà montré que le produit officiel n'est pas en mesure de reproduire les faibles valeurs d'humidité et a une très faible dynamique (Gruhler et al., 2008, Rudiger et al., 2009, Draper et al., 2009, Gruhier et al., 2010, Chaurasia et al., 2011). Parmi les autres produits d'humidité d'AMSR-E, l'algorithme de *Vrije Universiteit Amsterdam* (VUA, Owe et al. (2001)) retrouve l'humidité du sol et l'épaisseur optique de la végétation à partir des fréquences 6.9 et 10.7 GHz (cette dernière étant utilisée lorsque 6.9 GHz est inutilisable à cause des interférences) ainsi que la température de surface avec la fréquence 36.5 GHz. Les produits (L3 version 3 Grid) représentent les humidités instantanées et sont disponibles sur une grille $0.25^\circ \times 0.25^\circ$, uniquement pour les orbites descendantes (01:30). La qualité de ces données a été contrôlée au préalable et les estimations erronées dues à des reliefs trop importants (montagnes), à des conditions extrêmes (inondations ou neige) ont été retirées avant leur distribution.

A cette fréquence, on peut estimer que l'humidité retrouvée est représentative des deux premiers centimètres du sol. La Terre est entièrement observée tous les 2-3 jours.

3.1.2.2 L'instrument ASCAT

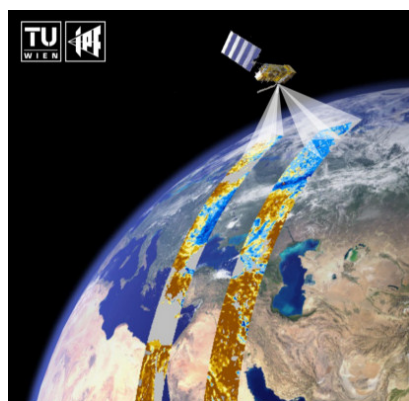


FIGURE 3.2 – L'instrument ASCAT.

l'orbite ascendante et de 09:30 (HSL) pour l'orbite descendante. L'algorithme développé à l'Université Technique de Vienne en Autriche (Wagner et al., 1999), fournit un indice d'humidité du sol qui varie entre 0 (sec) et 1 (saturé en eau) et qui n'a pas d'unité. Cet indice n'est pas linéaire



FIGURE 3.1 – Le satellite Aqua.

L'*Advanced Scatterometer* (ASCAT, Figa-Saldana et al. (2002)) a été lancé à bord du satellite MetOp-A en Octobre 2006. Cet instrument envoie des ondes électromagnétiques à une fréquence de 5.3 GHz et mesure l'onde électromagnétique réfléchie par la surface terrestre. Il rentre donc dans la catégorie des instruments actifs. Il est composé de six faisceaux, trois de chaque côté du satellite avec des angles azimutaux de 45° , 90° et 135° , avec des angles d'incidence de 25° à 64° et génère ainsi deux fauchées de 550 km de large chacune avec une résolution spatiale de 50 km. Les produits d'humidité sont échantillonnés avec une résolution de 25km et sont des produits instantanés. Les heures de passage à l'équateur sont de 21:30 (HSL) pour

car la valeur 1 ne correspond pas tout le temps à la même quantité volumique d'eau contenue dans le sol puisque le volume de saturation dépend majoritairement de la composition de ce sol (sable, argile). ASCAT fait suite au scatteromètre d'ERS 1 et 2 (*European Remote-Sensing Satellite*) lancés en 1991 et 1995 et les indices d'humidité d'ASCAT ont été préalablement calibrés entre le minimum et le maximum de l'humidité observée par la série ERS pour chaque pixel entre les années 1992 et 2007. Ils sont représentatifs des deux premiers centimètres du sol. L'intégralité de la surface terrestre est observée tous les 5 jours.

3.1.3 Données du modèle ECMWF

Le centre européen météorologique ECMWF (centre européen de prévisions météorologiques à moyen terme) produit de façon opérationnelle des prévisions météorologiques globales à moyen terme et dans ce procédé, il produit des cartes globales des variables environnementales qui ont permis d'arriver à ces prédictions. Parmi ces variables, on trouve l'humidité du sol (des 7 premiers centimètres sous la surface), la température du sol, la température de l'air ou encore l'épaisseur de la couche neigeuse.

L'algorithme SMOS utilise des données extérieures pour initialiser certaines variables du modèle et pour estimer la contribution au signal de certaines parties de la surface observée. Les données ECMWF sont utilisées comme données d'entrée de l'algorithme SMOS. Pour cela, un produit interne ECMWF a été créé pour représenter au mieux ces variables à la résolution et surtout à l'heure de passage du satellite SMOS (interpolations spatiales et temporelles). Dans cette étude, ce produit interne a été utilisé. Leur résolution native est de 0.225° toutes les trois heures.

3.2 Présentation des quatre bassins versants tests aux États-Unis

Basé sur les différents enjeux énumérés précédemment, un réseau de stations développé sur quatre bassins versants (ou zones géographiques) aux États-Unis a été conçu pour la validation des données AMSR-E sur les critères suivants (Jackson et al., 2010) :

- plusieurs stations sont positionnées à l'intérieur d'un *footprint* et doivent couvrir sa taille moyenne (25-50 km),
- les réseaux couvrent une large variété de végétations et de conditions climatiques,
- les instruments utilisés sont les mêmes pour toutes les stations,
- les mesures sont effectuées à 5 cm pour toutes les stations, et
- au moins une mesure par heure est effectuée (pour pouvoir être comparée ensuite à l'heure de passage du satellite la plus proche).

Dans ce cadre, quatre bassins versants aux États-Unis ont été sélectionnés et instrumentalisés par l'Agricultural Research Service (ARS) depuis 2002 : Walnut Gulch en Arizona, Little Washita dans l'Oklahoma, Little River en Géorgie et Reynolds Creek dans l'Idaho. Ils représentent différents types de climat (de semi-aride à humide) et différents types d'occupation du sol (cultures, forêts, ...). Le tableau 3.1 récapitule les caractéristiques de chaque bassin versant et la figure 3.3 montre leurs localisations et les occupations des sols.

Depuis 2002, des sondes d'humidité et de température ont acquis des données sur ces 4 bassins versants. Les données terrain, fournies par l'équipe en charge à l'USDA, sont des moyennes de toutes les sondes situées sur chaque bassin versant. Ce n'est pas une moyenne arithmétique mais une moyenne pondérée par des coefficients provenant de l'application d'un polygone de Thiessen qui permet de prendre en compte notamment les distances entre chaque station et de leur représentativité (Jackson et al., 2010). Ce jeu de données a été distribué pour la période

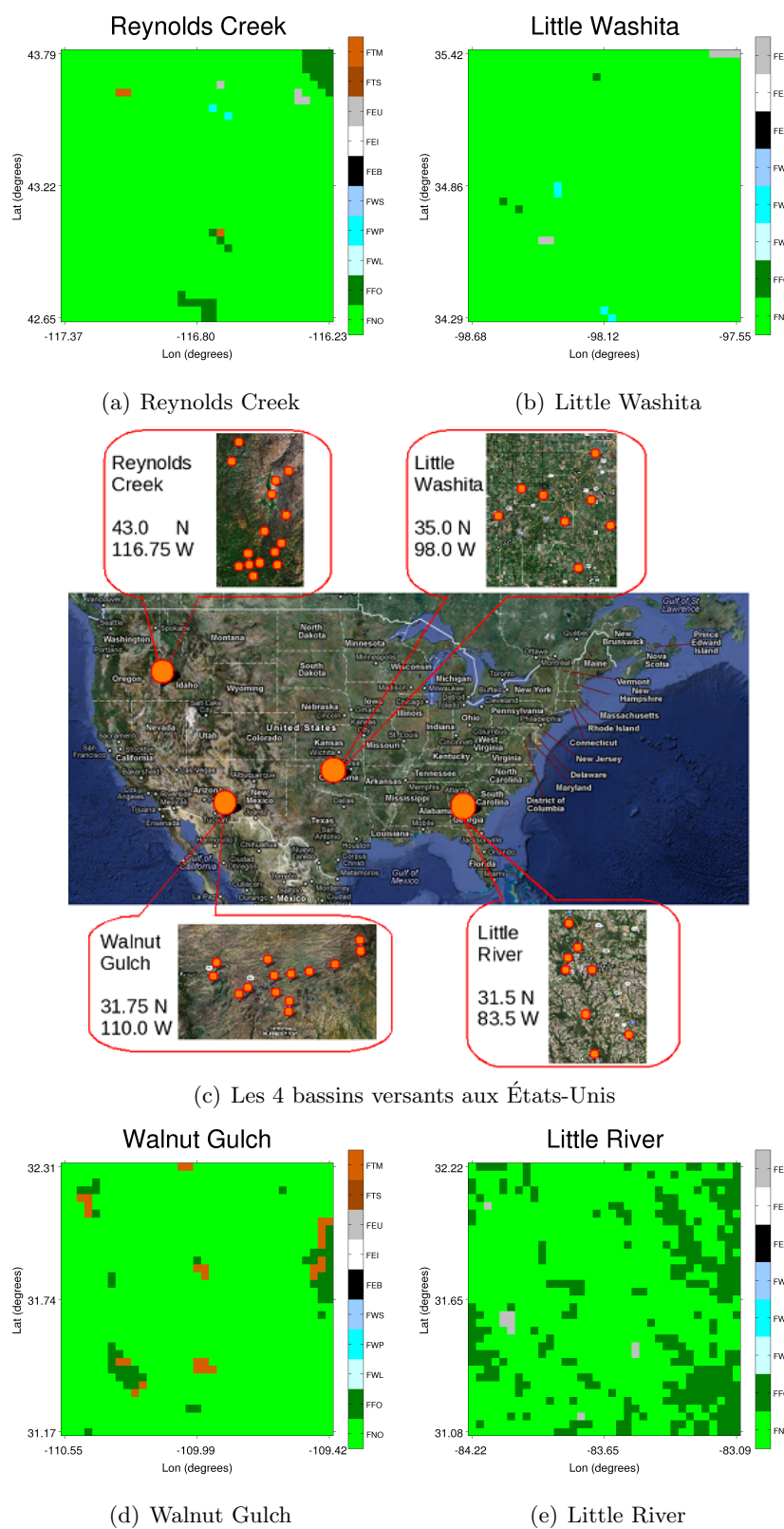


FIGURE 3.3 – Les quatre bassins versants tests pour comparer les données SMOS aux Etats-Unis. Pour chaque bassin versant, une représentation de l'occupation du sol est fournie par ECOCLIMAP (FTM : topographie modérée, FTS : topographie forte, FEU : zone urbaine, FEI : glace, FEB : zone aride, FWS : eau salée, FWP : eau douce, FWL : zone humide, FFO : forêt et FNO : zone de végétation basse)

Bassin versant	Nb. stations	Climat	Précip. annuelles (mm)	Topographie	Sol
Walnut Gulch	14	semi-aride	320	vallonné	pâturages
Little Washita	8	subhumide	750	vallonné	pâturages/blé
Little River	8	humide	1200	plat	cultures/forêts
Reynolds Creek	15	aride	500	montagneux	pâturages

Tableau 3.1 – Caractéristiques des quatre bassins versants : le nombre de stations qui mesurent l'humidité, le climat, les précipitations annuelles, la topographie et le type majoritaire d'utilisation du sol (Jackson et al., 2010).

2002-2011 et dû à une grande réorganisation du réseau en 2005 et 2006, seulement un nombre limité de sondes a été retenu. Même si dans ce chapitre seules les données 2010 sont utilisées, uniquement les données provenant des stations jugées fiables sur la période de temps 2002-2011 ont été fournies par l'équipe de l'USDA.

De plus, plusieurs sondes ont été écartées de la moyenne à cause de leur performance faible ou douteuse selon les critères suivants : (1) les sondes avec des indisponibilités temporelles n'ont pas été utilisées et seules les stations avec des mesures continues ont été prises en compte ; (2) les sondes ont été calibrées et vérifiées pendant une campagne de terrain et celles qui n'étaient pas en accord avec les mesures réalisées lors de cette campagne ont été éliminées ; (3) une analyse temporelle de stabilité a été réalisée avec toutes les mesures terrain et les sondes jugées trop différentes (moyenne trop élevée ou trop grande variabilité) ont été retirées ; (4) les sondes ne répondant pas correctement visuellement aux événements pluvieux n'ont pas été distribuées. Cette étape de filtrage des stations a été réalisée en amont par l'équipe de l'USDA (Jackson et al., 2010). L'humidité et la température du sol ont été mesurées toutes les 30 minutes sur tous les sites sauf sur Reynolds Creek où les mesures ont été effectuées toutes les heures.

3.2.1 Walnut Gulch



FIGURE 3.4 – Walnut Gulch, (Jackson et al., 2010).

Walnut Gulch (WG, fig. 1.4) est situé dans le Sud-Est de l'Arizona, état situé dans le Sud-Ouest des États-Unis. La majeure partie de ce bassin versant est recouverte d'arbustes et d'herbes du désert, ce qui est typique de cette région semi-aride. La température annuelle moyenne est de 17.6 °C et les précipitations annuelles s'élèvent à 320 mm (les deux tiers provenant des orages en fin d'été). Les premiers centimètres du sol sont constitués à 60 % de gravier. Le sol est majoritairement occupé par des pâturages.

3.2.2 Little Washita



FIGURE 3.5 – Little Washita, (Jackson et al., 2010).

Little Washita (LW, fig. 1.5) se situe dans le Sud-Ouest de l'état d'Oklahoma au Sud de la région des Grandes Plaines aux États-Unis. Doté d'un climat subhumide (climat entre aride et humide), les précipitations annuelles sont de 750 mm, principalement pendant les saisons du printemps et d'automne. Le relief est modérément vallonné avec un dénivelé maximum de 200 m. Le sol est majoritairement occupé par des cultures et des prés et présente de grandes disparités en terme de texture (sable, argile).

3.2.3 Little River

Little River (LR, fig. 1.6) est localisé dans le Sud de l'état de Géorgie, au Sud-Est des États-Unis. Avec des précipitations annuelles de 1200 mm, le climat est considéré comme humide. Ce bassin versant est typique de cette région appelée l'*Atlantic Coastal Plain*, où il existe très peu de relief et où l'océan se situe à moins de 100 km et où les sols peuvent devenir rapidement marécageux dû à cette proximité avec l'océan. Approximativement 35 % du bassin versant est recouvert de forêts, 40 % de cultures et 20 % de prés, le reste représentant des zones inondées (marécages) et des zones urbaines.



FIGURE 3.6 – Little River, (Jackson et al., 2010).

3.2.4 Reynolds Creek



FIGURE 3.7 – Reynolds Creek, (Jackson et al., 2010).

Reynolds Creek (RC, fig. 1.7) est situé dans une région montagneuse, appelée les Rocheuses, au Sud-Ouest de l'Idaho, état dans le Nord-Ouest des États-Unis. Le relief de ce bassin versant y est élevé avec un dénivelé maximum de 1000 m, engendrant ainsi plusieurs types de climat, de sol et de végétation. Le climat est considéré comme semi-aride avec des précipitations annuelles de 500 mm. Dans la partie située en haute altitude, 75 % de ces précipitations sont en réalité des chutes de neige alors que la neige ne représente que 25 % à basse altitude.

3.3 Évaluation des données SMOS avec les données terrain et d'autres produits d'humidité

3.3.1 Présentation des données

Les produits d'humidité SMOS de niveau 2 version 4 sont comparés aux humidités VUA d'AMSR-E de niveau 3 version 3 *Grid*, aux indices d'humidité ASCAT de niveau 2 distribués par Eumetsat à 25 km et aux humidités du modèle ECMWF, produit de prévision interne à l'algorithme SMOS interpolé spatialement et temporellement.

Dans cette étude, des informations supplémentaires concernant la végétation et la présence de neige ont été nécessaires à l'analyse. En tant qu'indicateur de la quantité de végétation présente sur chaque bassin versant, l'indice de surface foliaire (ou LAI) de ECOCLIMAP a été utilisé. Le LAI est défini comme la moitié de la surface foliaire totale par unité de surface au sol (Chen & Black, 1992), prenant une valeur nulle pour un sol entièrement nu et pouvant dépasser parfois 10 pour des forêts très denses. Ce produit est disponible tous les 10 jours avec une résolution native de 1 km. L'indice LAI de ECOCLIMAP est un produit auxiliaire dans l'algorithme SMOS et il a été agrégé à 4 km pour correspondre à la résolution de la surface de travail (définie précédemment en section 2.2.3, Kerr et al. (2012)). C'est ce produit auxiliaire qui a été utilisé dans cette étude. Concernant la végétation, puisqu'aucune mesure sur le terrain n'a été réalisée, une moyenne des LAI des pixels de chaque bassin versant a été utilisée comme référence.

La variable ECMWF de la couverture neigeuse a également été utilisée afin d'identifier les jours où il y a eu des événements neigeux. La couverture neigeuse représente la quantité de neige en cm présente à la surface du sol.

3.3.2 Pré-traitement des données

Dans un premier temps, les produits VUA et ASCAT ont été interpolés sur la grille SMOS puis dans un second temps, tous les produits ont été moyennés suivant le même principe que les moyennes de mesures terrain par station. Ainsi toutes les moyennes peuvent être équitablement comparées. Toujours dans cet esprit d'égalité, les comparaisons ont été effectuées uniquement sur les jours en commun qui ne comportaient aucun événement de neige ou de gel pour les 5 jeux de données : SMOS, ECMWF, VUA, ASCAT et les mesures terrain. A cause d'un changement dans le modèle ECMWF en Novembre 2010, les statistiques ne tiennent pas compte des données après cette date et ne seront donc représentatives que de la période 1er Janvier - 9 Novembre 2010.

Après cette étape de filtrage, le nombre de jours en commun disponibles est de 48/57/44/50 pour la comparaison du matin sur les bassins WG/LW/LR/RC et de 61/72/50/70 pour l'après-midi. Plus de dates sont disponibles dans l'après-midi puisque le jeu de données VUA n'est disponible que le matin, ce qui retire une contrainte supplémentaire d'un produit satellite pour l'après-midi. Les événements neigeux ont été identifiés grâce à la variable ECMWF : quand la couverture neigeuse était supérieure à 0 cm, les données de ce jour n'étaient pas considérées. Les événements de gel ont été détectés avec la température du sol mesurée sur chaque bassin versant : lorsqu'elle était inférieure à -3°C , le sol était considéré gelé et les données n'ont pas été utilisées.

3.3.3 Méthodologie

Les comparaisons ont été effectuées sur l'année 2010, couvrant ainsi les 4 saisons. Le coefficient de corrélation (R), l'erreur standard d'estimation entre les produits d'humidité des sols

et les données terrain (SEE), le biais et l'erreur globale (RMSE) ont été calculés pour chaque bassin versant avec les mesures terrain les plus proches des temps de passage des satellites (6h/18h pour SMOS et ECMWF, 1h30 pour VUA et 10h/20h pour ASCAT). La corrélation mesure l'intensité d'une possible relation linéaire entre deux jeux de données. La SEE mesure la différence de variabilité entre deux jeux de données, si $SEE=0$ alors les deux jeux ont la même variabilité sans prendre en compte les possibles biais existants. Le biais est la différence des moyennes des deux jeux de données testés. Le RMSE prend en compte le biais et la différence en variabilité.

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad (3.1)$$

$$SEE^2 = \frac{1}{N} \sum ((x_i - y_i) - (\bar{x} - \bar{y}))^2 \quad (3.2)$$

$$\text{biais}^2 = (\bar{x} - \bar{y})^2 \quad (3.3)$$

$$RMSE^2 = \frac{\sum (x_i - y_i)^2}{N} \quad (3.4)$$

$$RMSE^2 = SEE^2 + \text{biais}^2 \quad (3.5)$$

Étant donné le faible nombre de jours disponibles communs à tous les produits d'humidité (entre 44 et 72) pour un calcul statistique, des intervalles de confiance à 95% ont été calculés en complément des coefficients de corrélation R et des erreurs SEE . En effet, plus le nombre de points est faible, plus l'intervalle de confiance sera large et moins il sera facile de conclure. Le détail théorique du calcul des intervalles de confiance est présenté en annexe A.1. Ces statistiques s'accompagnent également d'un paramètre *valeur p* qui représente la probabilité de commettre une erreur. Si cette valeur p est supérieure à un seuil de 5% alors on dit que la mesure statistique associée n'est pas significative et elle ne sera pas prise en compte dans l'analyse.

Le produit ASCAT est un indice d'humidité variant entre 0 et 1 et ne donne pas de valeur volumétrique de l'humidité dans le sol. Pour cela, il doit être traité séparément des autres produits d'humidité. Il existe plusieurs choix pour le rendre comparable aux autres jeux de données : (1) redimensionner le produit ASCAT linéairement entre le minimum et le maximum des données terrain ; (2) redimensionner le produit ASCAT statistiquement par CDF (*Cumulative Density Function*) matching avec les données terrain (méthode statistique qui fait correspondre les distributions statistiques de deux jeux de données). Ces deux approches sont abordées dans les sections suivantes.

3.3.4 Comparaison avec redimensionnement linéaire du produit ASCAT

3.3.4.1 Walnut Gulch

Walnut Gulch (WG) est le site le plus sec de cette étude. Les mesures d'humidité ne dépassent que rarement les $0.10 \text{ m}^3/\text{m}^3$ (fig. 3.8). Concernant les statistiques du matin, SMOS a obtenu les meilleurs résultats avec un très bon coefficient de corrélation ($R=0.90$, tab. 3.2 et fig. 3.13) associé au biais le plus bas ($0.015 \text{ m}^3/\text{m}^3$) et à l'erreur globale la plus faible ($0.048 \text{ m}^3/\text{m}^3$) par rapport aux autres jeux de données testés. Pour les statistiques de l'après-midi, SMOS et ECMWF sont très proches et significativement meilleurs que ASCAT en terme de corrélation (0.73 et 0.75) mais SMOS présente un biais bien plus faible ($0.017 \text{ m}^3/\text{m}^3$ contre $0.165 \text{ m}^3/\text{m}^3$ pour ECMWF). Sur ce bassin versant, le produit d'humidité SMOS divise le biais d'un facteur 8 et l'erreur globale d'un facteur 3 comparé aux produits ECMWF, VUA et ASCAT tout en gardant un niveau de corrélation équivalent voire meilleur.

		AM				PM					
		R	SEE	biais	RMSE	#pts	R	SEE	biais	RMSE	#pts
WG	SMOS	0.90	0.046	0.015	0.048	48	0.73	0.054	0.017	0.056	61
	ECMWF	0.69	0.031	0.169	0.172		0.75	0.031	0.165	0.168	
	VUA	0.79	0.063	0.128	0.141		-	-	-	-	
	ASCATn	0.59	-	-	-		0.42	-	-	-	
LW	SMOS	0.80	0.047	0.014	0.049	57	0.62	0.061	0.001	0.061	72
	ECMWF	0.84	0.042	0.110	0.118		0.71	0.049	0.099	0.111	
	VUA	0.72	0.062	0.113	0.129		-	-	-	-	
	ASCATn	0.73	-	-	-		0.60	-	-	-	
LR	SMOS	0.70	0.038	0.042	0.057	44	0.64	0.066	0.091	0.112	50
	ECMWF	0.68	0.046	0.255	0.259		0.84	0.040	0.267	0.270	
	VUA	0.66	0.053	0.156	0.163		-	-	-	-	
	ASCATn	0.22*	-	-	-		0.36	-	-	-	
RC	SMOS	0.79	0.039	-0.038	0.055	50	0.61	0.061	-0.012	0.061	70
	ECMWF	0.89	0.030	0.071	0.077		0.81	0.035	0.060	0.069	
	VUA	0.83	0.066	0.061	0.087		-	-	-	-	
	ASCATn	0.70	-	-	-		0.46	-	-	-	

Tableau 3.2 – Résultats statistiques entre les produits d'humidité (satellites et modèles) et les mesures *in situ* sur les 4 bassins versants tests avec AM correspondant aux orbites ascendantes du matin pour SMOS (6:00), ECMWF (6:00), AMSR-E (1:30) et descendantes du matin pour ASCAT (9:30). Les données ASCAT ont été normalisées linéairement dans la gamme des valeurs in situ (appelées ASCATn). Les meilleurs résultats sont indiqués en gras. (*) dénote un résultat statistique non significatif qui correspond à une valeur p supérieure à 0.05.

En s'intéressant de plus près à la tendance saisonnière, on remarque que les données SMOS tendent à sous-estimer l'humidité pendant la saison sèche (Mars-Juillet) et donnent de meilleures performances pendant l'hiver (Octobre-Décembre). Cependant, tous les événements pluvieux ont été correctement identifiés et retranscrits dans la série temporelle de SMOS. La version 4 des produits d'humidité SMOS utilisée dans cette étude a été produite avec des étalonnages différents de températures de brillance acquises, ce qui pourrait expliquer une amélioration du biais au fil de l'année 2010. Cet artefact devrait disparaître avec la version 5 qui utilisera un étalonnage unique des températures de brillance. Cependant, les températures de brillance avant Avril 2010 resteront d'une qualité moyenne puisque l'instrument a subi des changements notables durant la phase de recette en vol de Novembre 2009 à Avril 2010.

Le produit ECMWF obtient le biais le plus fort avec les données terrain (fig. 3.8 et fig. 3.12). Néanmoins, ECMWF est plus proche des mesures d'humidité après mi-Novembre quand le changement de système de modélisation a été opéré et qui avait pour but de diminuer les biais sur les surfaces sèches. Après mi-Novembre, le biais ECMWF devient plus faible que celui de VUA. Le produit ECMWF suit correctement la dynamique de l'humidité du sol. La majeure partie du biais d'ECMWF provient de ce niveau de saturation bas de l'humidité dont le niveau se situe autour de $0.20 \text{ m}^3/\text{m}^3$ avant la mi-Novembre puis autour de $0.12 \text{ m}^3/\text{m}^3$ après.

D'après la fig. 3.8, les données VUA sont plus proches des données terrain en tout début d'été, en Juin, lorsque le sol est très sec (moins de $0.1 \text{ m}^3/\text{m}^3$), que la température est très élevée et qu'il n'y a pas eu de pluie depuis plusieurs semaines. Le biais VUA le plus élevé ($0.15 \text{ m}^3/\text{m}^3$) se situe en hiver, Octobre-Février, lorsque la végétation et la température sont au plus bas.

La série temporelle d'ASCAT se révèle être très bruitée et même si les événements de pluie sont apparents (les valeurs les plus hautes sur la fig. 3.12), la saison sèche n'est pas bien décrite quand l'humidité du sol est basse et stable. Les observations micro-ondes actives sont très sensibles à la rugosité de surface et Walnut Gulch est composé en majorité de sols sableux avec une quantité importante de graviers. Ces caractéristiques du sol font que la rugosité est importante et peut engendrer de mauvais résultats quand on regarde avec un système micro-onde actif (une observation dans l'actif étant beaucoup plus sensible à la rugosité de surface que dans le passif).

3.3.4.2 Little Washita

Le bassin versant de Little Washita (LW) a la plus large gamme de valeurs mesurées d'humidité (de 0.1 à $0.3 \text{ m}^3/\text{m}^3$, fig. 3.9). Plusieurs jours ont dû être écartés de l'analyse en Janvier-Février à cause des événements neigeux et de gel du sol. Concernant les résultats du matin, SMOS et ECMWF ont obtenu de meilleurs coefficients de corrélation que VUA et ASCAT mais ne sont pas significativement meilleurs car leurs intervalles de confiance se chevauchent, donc on ne peut pas affirmer clairement quel jeu de données a obtenu la meilleure corrélation avec les données mesurées sur le terrain. Avec un biais de $0.014 \text{ m}^3/\text{m}^3$ (tab. 3.2 et fig. 3.13) et une erreur globale de $0.049 \text{ m}^3/\text{m}^3$, le produit SMOS est le plus proche des données terrain. Concernant l'après-midi, aucun produit ne peut être identifié comme le meilleur en terme de corrélation (0.62 pour SMOS, 0.71 pour ECMWF et 0.60 pour ASCAT). Cependant, SMOS ne présente quasiment aucun biais ($0.001 \text{ m}^3/\text{m}^3$). Pour ce bassin versant, SMOS a divisé le biais d'un facteur 8 et l'erreur globale d'un facteur 2 par rapport aux autres jeux de données pour une corrélation équivalente.

La figure 3.9 montre que les produits SMOS, VUA et ECMWF décrivent correctement la dynamique de l'humidité répondant bien aux différents événements pluvieux, sauf au début de l'année (Janvier-Avril) lorsque des pluies faibles et consécutives se sont produites impliquant des valeurs d'humidité plus dispersées pour SMOS et VUA. Les pluies de Mai, Juin et Juillet

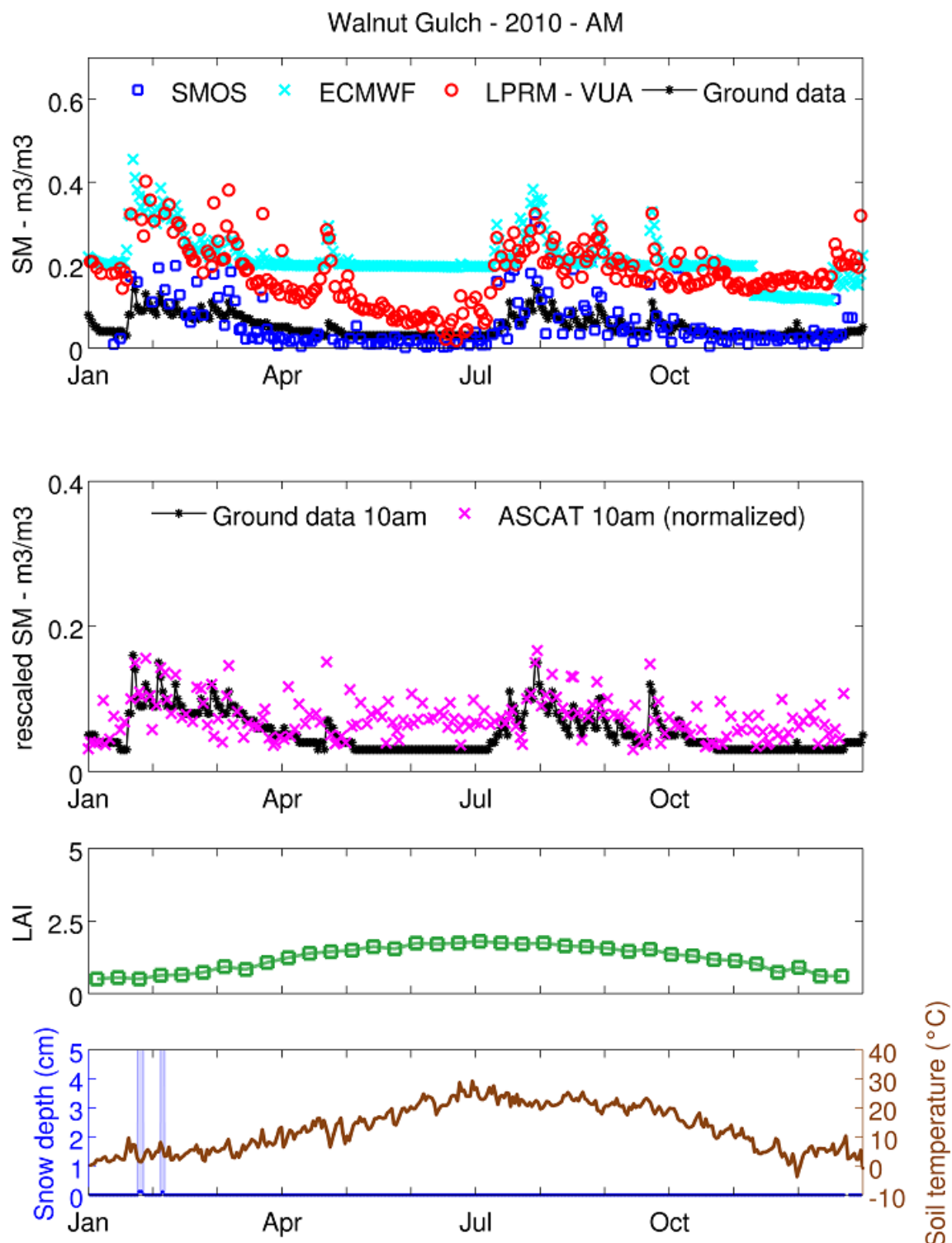


FIGURE 3.8 – Série temporelle de l'humidité du sol (SM) du matin sur Walnut Gulch retrouvée par SMOS, ECMWF, VUA et les données terrain (1er panneau) ; humidité du sol linéairement normalisée d'ASCAT (*rescaled SM*) et les données terrain (2ème panneau) ; les données de végétation LAI d'ECOCLIMAP ainsi que l'épaisseur de neige d'ECMWF avec la température mesurée au sol (*snow depth* et *soil temperature*). Les jours de neige et de gel sont indiqués en ombres bleues sur le panneau du bas.

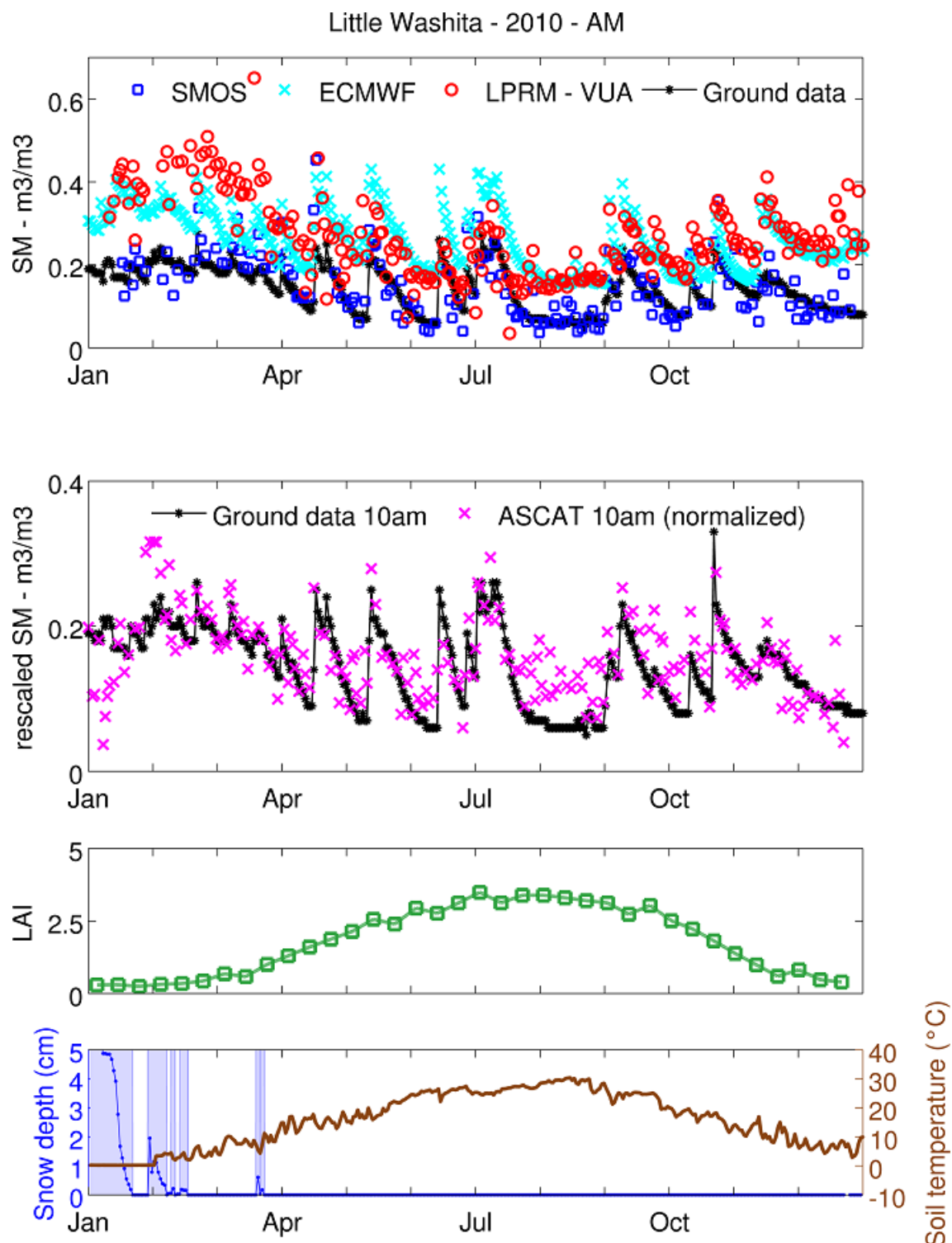


FIGURE 3.9 – Série temporelle de l'humidité du sol (SM) du matin sur Little Washita vue par SMOS, ECMWF, VUA et les données terrain (1er panneau) ; humidité du sol linéairement normalisés d'ASCAT ($rescaled\ SM$) et les données terrain (2ème panneau) ; les données de végétation LAI d'ECOCLIMAP ainsi que l'épaisseur de neige d'ECMWF avec la température mesurée au sol ($snow\ depth$ et $soil\ temperature$). Les jours de neige et de gel sont indiqués en ombres bleues sur le panneau du bas.

sont particulièrement bien reproduites par SMOS et ECMWF avec des phases de ressuyage clairement identifiables. Cependant, on peut noter que très souvent, la première valeur d'humidité retrouvée par SMOS après un événement pluvieux est sur-estimée. La bande L étant très sensible à l'eau et il est possible que lorsqu'il pleut, des flaques se forment sur le sol et que l'humidité retrouvée par SMOS soit ainsi sur-estimée.

Le produit ECMWF présente de très bons résultats en corrélation (0.84 pour le matin et 0.71 pour l'après-midi, tab. 3.2) mais il se retrouve pénalisé par son biais tout au long de l'année (environ $0.10 \text{ m}^3/\text{m}^3$) qui peut être clairement identifié sur la figure 3.12 et qui provoque de fortes valeurs de RMSE.

Le produit VUA montre un biais variable avec les saisons : plus de $0.2 \text{ m}^3/\text{m}^3$ en Janvier-Février et quasi nul en Juin-Juillet. Comme observé pour Walnut Gulch, le biais le plus faible est observé lorsque la végétation et la température de surface sont à leur maximum. Pendant cette période d'été, les données VUA étaient plus proches des mesures terrain que les données ECMWF. Ce biais non stable empêche le produit VUA d'avoir une bonne corrélation sur l'année. Sans ce biais, la dynamique de l'humidité du sol avec les différents événements pluvieux est respectée.

Comme remarqué sur Walnut Gulch, l'indice d'humidité d'ASCAT est encore très bruité sur Little Washita. Cependant le produit ASCAT s'améliore sur les périodes Février-Juin et Novembre-Décembre. Pendant la saison sèche d'été, ASCAT montre un biais en Juillet-Août (fig. 3.9).

3.3.4.3 Little River

Little River (LR) est le site avec la plus haute fréquence d'événements pluvieux n'impliquant ainsi que de très faibles variations de l'humidité du sol (fig. 3.10). Concernant les statistiques du matin, tous les produits sont équivalents en terme de corrélation (autour de 0.68), sauf pour ASCAT dont la corrélation est de 0.22. Cependant, cette valeur n'est pas statistiquement significative car sa valeur p associée (0.15) est supérieure au niveau de confiance de 0.05 (ou 5%). Pour cette raison, cette statistique n'est pas montrée sur la fig. 3.13. Le produit SMOS a cependant le biais le plus faible ($0.042 \text{ m}^3/\text{m}^3$) et l'erreur globale la plus faible ($0.057 \text{ m}^3/\text{m}^3$). Pour l'après-midi, même si l'intervalle de confiance du coefficient de corrélation de SMOS chevauche légèrement celui d'ECMWF, on peut supposer que le produit ECMWF a la meilleure corrélation avec une valeur de 0.84. SMOS reste tout de même le produit le plus proche des données avec un biais de $0.091 \text{ m}^3/\text{m}^3$ et une erreur globale de $0.112 \text{ m}^3/\text{m}^3$. Sur ce bassin versant, SMOS a divisé le biais d'un facteur 3 et l'erreur globale d'un facteur 2 mais n'a cependant pas réussi à garder un niveau de corrélation équivalent au produit ECMWF.

En observant l'évolution de la série temporelle de l'humidité des sols sur la fig. 3.10, on peut voir que les petits événements pluvieux successifs de la première moitié de l'année ont été observés avec beaucoup de difficulté par les produits satellite (SMOS, VUA et ASCAT), ce qui a eu pour conséquence une dispersion des valeurs d'humidité de ces produits autour des mesures terrain. A partir de Septembre, lorsque les pluies se sont faites plus espacées, SMOS se rapproche des mesures in situ mais Little River est le bassin versant où SMOS est le plus éloigné de ces objectifs de mission en terme d'erreur. Pour rappel, un des objectifs de la mission SMOS est de mesurer l'humidité des sols avec une précision de $0.04 \text{ m}^3/\text{m}^3$. Cependant, Little River est le site le plus hétérogène en terme d'occupation des sols avec beaucoup de forêts et de zones humides dont les contributions au signal doivent être estimées afin de ne retrouver l'humidité que sur les zones à faible végétation. Cette étape est très délicate puisqu'elle fait intervenir des données auxiliaires extérieures qui ajoutent des incertitudes non négligeables.

Le produit ECMWF présente encore un biais haut et stable tout au long de l'année ($0.26 \text{ m}^3/\text{m}^3$) mais il continue à très bien modéliser la dynamique de l'humidité, ainsi que tous les

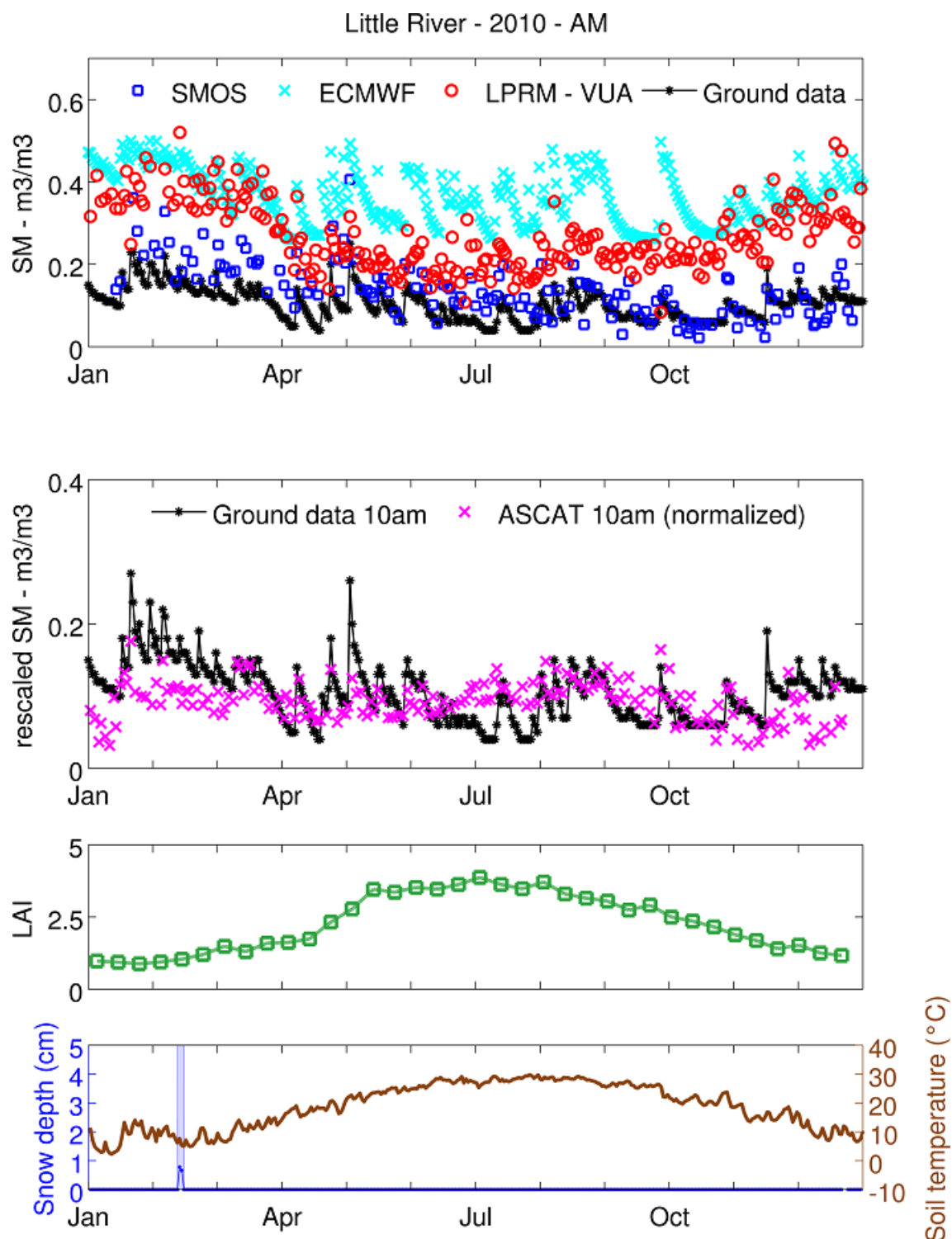


FIGURE 3.10 – Série temporelle de l'humidité du sol (SM) du matin sur Little River vue par SMOS, ECMWF, VUA et les données terrain (1er panneau); humidité du sol linéairement normalisés d'ASCAT ($rescaled\ SM$) et les données terrain (2ème panneau); les données de végétation LAI d'ECOCLIMAP ainsi que l'épaisseur de neige d'ECMWF avec la température mesurée au sol ($snow\ depth$ et $soil\ temperature$). Les jours de neige et de gel sont indiqués en ombres bleues sur le panneau du bas.

événements pluvieux contrairement aux produits satellite. Néanmoins, les gammes de valeur d'humidité décrites pendant ces pluies sont beaucoup trop larges comparées aux données terrain : $0.20 \text{ m}^3/\text{m}^3$ pour ECMWF contre $0.05 \text{ m}^3/\text{m}^3$ pour les mesures terrain. Les mois de Septembre et Octobre en sont de très bons exemples.

Le produit VUA présente également un biais très fort sur Little River : de 0.16 à $0.30 \text{ m}^3/\text{m}^3$. Comme pour les précédents bassins versants, le biais est plus faible en été lorsque la végétation et la température ont atteint leur maximum.

Les humidités d'ASCAT reflètent relativement bien la dynamique de l'humidité pour les mois de Mars à Juin mais plus du tout après Juin où les données deviennent bruitées et ne suivent plus l'évolution des données terrain. En effet, les mesures terrain diminuent en Juin et Juillet alors que les données ASCAT augmentent. Ceci est représenté par un large nuage de points sur la fig. 3.12 quand les données ASCAT sont comparées aux mesures terrain. La valeur de corrélation n'est donc pas statistiquement représentative de ce nuage de points.

3.3.4.4 Reynolds Creek

Le bassin versant de Reynolds Creek est situé dans une région montagneuse des Rocheuses. Ce site représente un ensemble de conditions très difficiles pour les satellites à cause de sa topographie (qui peut altérer la reconstruction de l'image et l'inversion, Mialon et al. (2008)) mais également à cause des nombreuses chutes de neige et des phénomènes de gel du sol très fréquents (fig. 3.11). Concernant les résultats de la comparaison du matin, tous les produits sont équivalents en terme de corrélation et de biais absolus : SMOS sous-estime l'humidité alors que ECMWF et ASCAT la sur-estime. Les valeurs d'erreur globale sont très basses pour ECMWF et VUA qui réalisent ici leurs meilleures performances en terme de RMSE ($0.077 \text{ m}^3/\text{m}^3$ et $0.087 \text{ m}^3/\text{m}^3$ respectivement). Pour les résultats de l'après-midi, la corrélation d'ECMWF est la meilleure (0.81) comparée à SMOS et ASCAT. SMOS a cependant le biais le plus faible ($0.012 \text{ m}^3/\text{m}^3$) mais obtient une erreur globale comparable à celle d'ECMWF (autour de $0.065 \text{ m}^3/\text{m}^3$). A l'exception du début de l'année, l'évolution temporelle générale de tous les produits est assez satisfaisante compte tenu de la complexité de ce bassin versant.

Reynolds Creek est le seul site où SMOS sous-estime l'humidité mesurée sur le terrain (fig. 3.11). Cependant, si on regarde la période sans neige ni gel du sol (Avril-Octobre), SMOS semble retrouver l'humidité proprement avec un biais bien moindre que pendant l'hiver. Cela peut être expliqué par le fait que, lorsque le sol est gelé, la constante diélectrique diminue et le sol apparaît comme étant sec (Matzler, 1994). La neige sèche est quant à elle, supposée être transparente en bande L (Armstrong, 1999).

Le biais d'ECMWF est plus faible sur Reynolds Creek que sur les autres bassins versants, et plus spécialement pendant le printemps et l'automne. Cependant, pendant l'été, ECMWF semble atteindre un niveau d'humidité minimum puisque ses valeurs se stabilisent autour de $0.2 \text{ m}^3/\text{m}^3$ alors que les données terrain indiquent $0.05 \text{ m}^3/\text{m}^3$ (fig. 3.11 et 3.12).

L'évolution de l'humidité VUA se révèle être très bruitée en Mai-Juin, ce qui pourrait être dû à la croissance de la végétation car les bandes C et X y sont très sensibles, beaucoup plus qu'en bande L. Mais entre Juin et Octobre, lorsque la végétation se stabilise et que la température est la plus élevée, le produit VUA retranscrit très bien la dynamique des mesures d'humidité avec un biais quasi nul.

Après le dernier événement neigeux en Mars, le produit ASCAT suit plutôt correctement l'évolution de l'humidité et obtient un bon coefficient de corrélation (0.70).

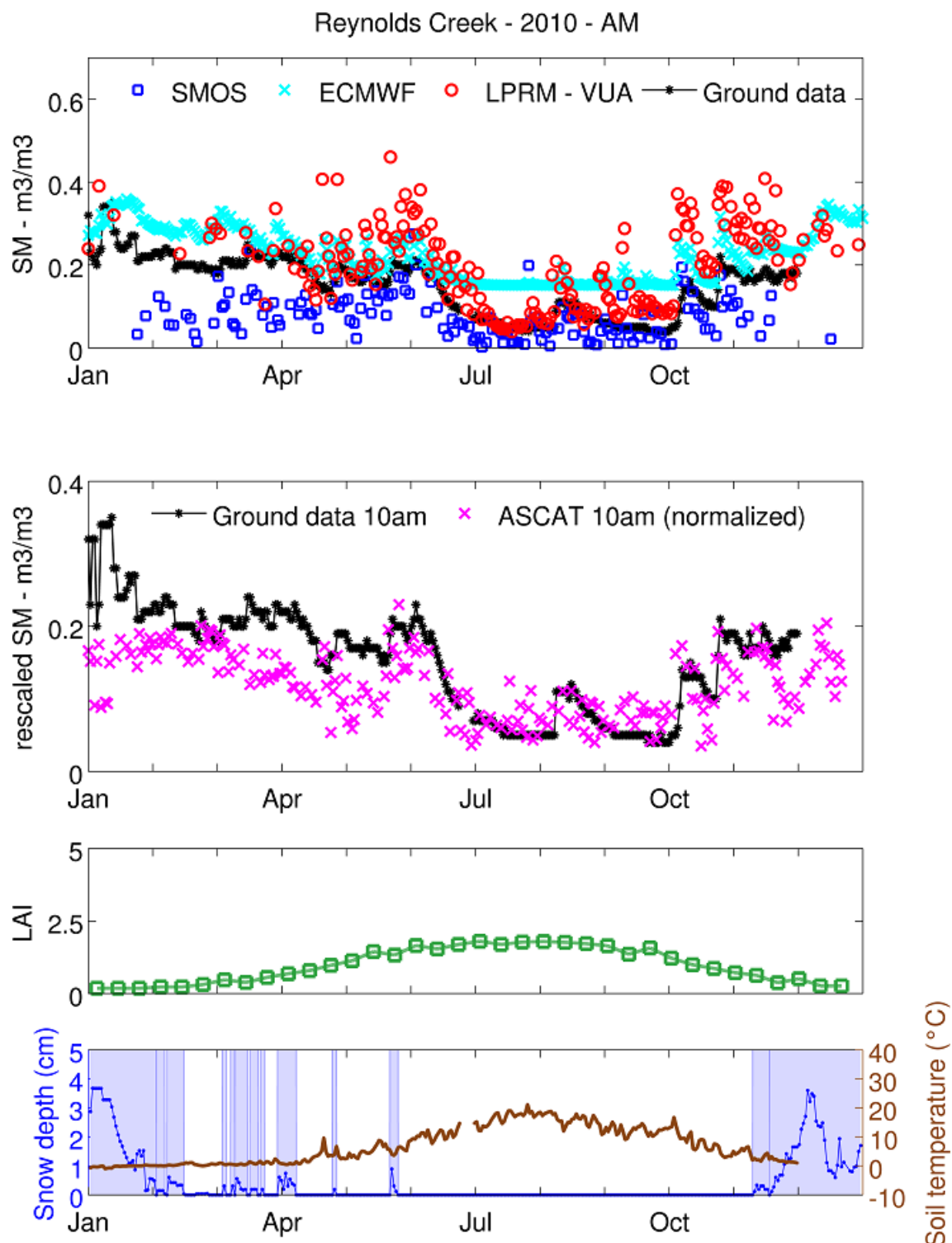


FIGURE 3.11 – Série temporelle de l'humidité du sol (SM) du matin sur Reynolds Creek vue par SMOS, ECMWF, VUA et les données terrain (1er panneau) ; humidité du sol linéairement normalisés d'ASCAT ($rescaled\ SM$) et les données terrain (2ème panneau) ; les données de végétation LAI d'ECOCLIMAP ainsi que l'épaisseur de neige d'ECMWF avec la température mesurée au sol ($snow\ depth$ et $soil\ temperature$). Les jours de neige et de gel sont indiqués en ombres bleues sur le panneau du bas.

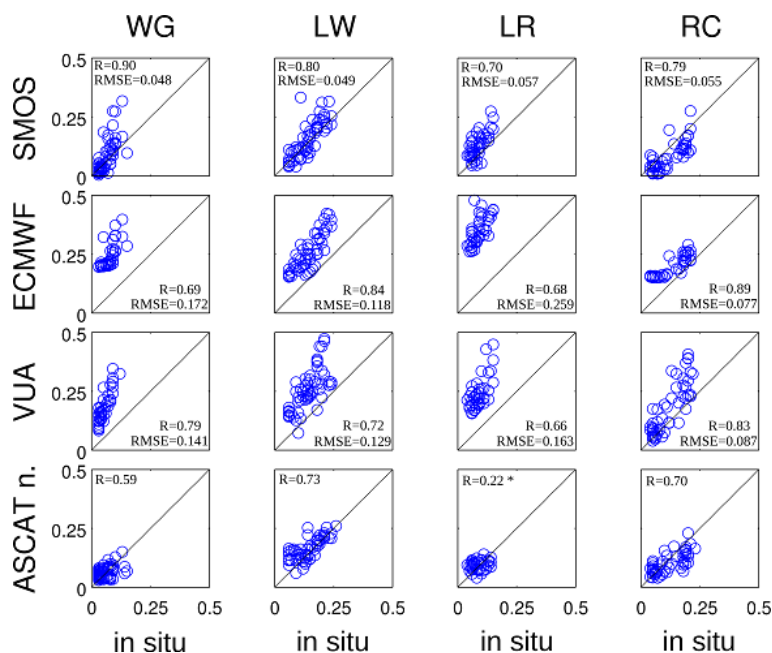


FIGURE 3.12 – Comparaison des humidités des sols retrouvées par SMOS, ECMWF, VUA et ASCAT normalisé avec les données d’humidité mesurées sur les quatre bassins versants en 2010 (matin uniquement). (*) dénote un résultat statistique non significatif avec une valeur p supérieure à 0.05.

3.3.4.5 Différences entre les résultats du matin et de l’après-midi

Les performances statistiques de l’après-midi (tab. 3.2 et fig. 3.13) montrent une moins bonne corrélation et une plus grande dispersion que le matin pour les produits satellite SMOS et ASCAT. Le coefficient de corrélation de SMOS est en moyenne de 0.80 pour le matin pour une erreur d’estimation SEE de $0.043 \text{ m}^3/\text{m}^3$ alors que la moyenne de l’après-midi pour la corrélation est de 0.65 pour une erreur de $0.061 \text{ m}^3/\text{m}^3$. Concernant ASCAT, sa corrélation du matin est en moyenne de 0.67 contre 0.49 pour l’après-midi (les résultats de Little River n’ont pas été pris en compte car ils n’étaient pas statistiquement significatifs). Il existe plusieurs explications possibles à cette différence de performance.

Le signal reçu par l’instrument SMOS est supposé provenir des 2 à 5 premiers centimètres du sol, largement conditionné par les propriétés du sol et les conditions d’humidité (plus profond pour un sol sableux et sec, plus en surface pour un sol argileux et humide). A 6:00h, lorsque le sol est en quasi-équilibre thermique, les observations de SMOS peuvent être comparées avec une plus grande confiance aux mesures terrain d’humidité à 5 cm. A 18:00h, le sol n’est plus en équilibre et a emmagasiné beaucoup de chaleur durant la journée et si SMOS ne reçoit que du signal des 3 premiers centimètres, les différences avec les mesures terrain à 5 cm seront bien plus grandes à 18:00h qu’à 6:00h. Jackson (1980) recommande en effet d’utiliser des mesures terrain du matin lorsqu’elles sont comparées à des observations satellites pour éviter ce découplage pendant la journée à cause du sol qui n’est plus en équilibre. Ces mêmes arguments sont également valables pour les résultats d’ASCAT à 10:00 et 22:00. Ces résultats sont d’ailleurs en accord avec ceux Albergel et al..

Une autre explication possible pour ces différences entre les résultats du matin et de l’après-midi est liée à la modélisation de la température de surface dans les algorithmes. L’hypothèse générale est de supposer que la température de la végétation est égale à la température du sol. Cette affirmation est plutôt vraie la nuit et le matin mais pendant la journée, la température

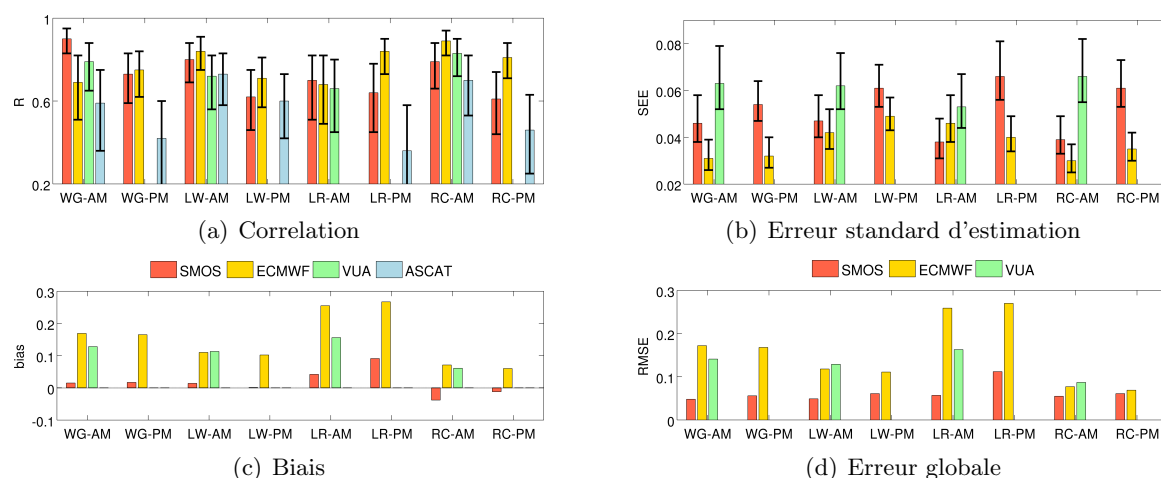


FIGURE 3.13 – Statistiques sur les comparaisons entre les produits d'humidité des sols de SMOS, ECMWF, VUA et les produits ASCAT normalisés linéairement avec les humidités mesurées sur les quatre bassins versants tests. Les 4 sous figures correspondent au coefficient de corrélation R (a), l'erreur d'estimation SEE (b), le biais (c) et l'erreur globale RMSE (d). Les données du matin (AM) et de l'après-midi (PM) ont été traitées séparément et les intervalles de confiance à 95% sont indiqués pour R et SEE.

	Walnut Gulch	Little Washita	Little River	Reynolds Creek
Matin	0.09%	0.00%	0.38%	2.23%
Après-midi	0.37%	0.38%	0.82%	0.36%

Tableau 3.3 – Probabilité d'occurrence de RFI (*Radio Frequency Interference*) en 2010 sur les quatre bassins versants test le matin et l'après-midi (ATBD, 2010)

de la végétation est censée être plus proche de la température de l'air que de la température du sol, plus spécialement lorsque les températures sont très élevées ou très faibles. Cette hypothèse de températures de végétation et de sol égales peut ainsi mener à des humidités retrouvées à partir de données satellite moins précises l'après-midi que le matin.

Concernant les résultats de SMOS plus spécifiquement, une dernière possibilité pourrait être la présence de RFI (*Radio Frequency Interference*) qui sont des interférences émises dans la bande L par l'homme au sol (télé-transmission, radio, communications,...). A partir des cartes de probabilité de contamination, qui ont été calculées à partir des polarisations croisées (ATBD, 2010), il est possible de calculer les probabilités d'occurrence des RFI sur chaque bassin versant le matin et l'après-midi (tab. 3.3). A l'exception de Reynolds Creek, les probabilités de contamination RFI sont plus élevées l'après-midi que le matin même si elles restent très faibles en comparaison d'autres régions comme l'Asie où plus de la moitié des acquisitions sont contaminées. La différence de probabilité de RFI entre le matin et l'après-midi peut expliquer ainsi partiellement les différences de performance de SMOS.

3.3.5 Comparaison par redimensionnement statistique du produit ASCAT

Dans la section précédente, les données ASCAT ont été linéairement redimensionnées entre les minimum et maximum des mesures terrain. Cette méthode peut paraître arbitraire du fait de sa linéarité entre deux extrêmes. Dans cette partie, un redimensionnement par une méthode statistique est proposée. Il s'agit alors de faire correspondre les deux distributions statistiques

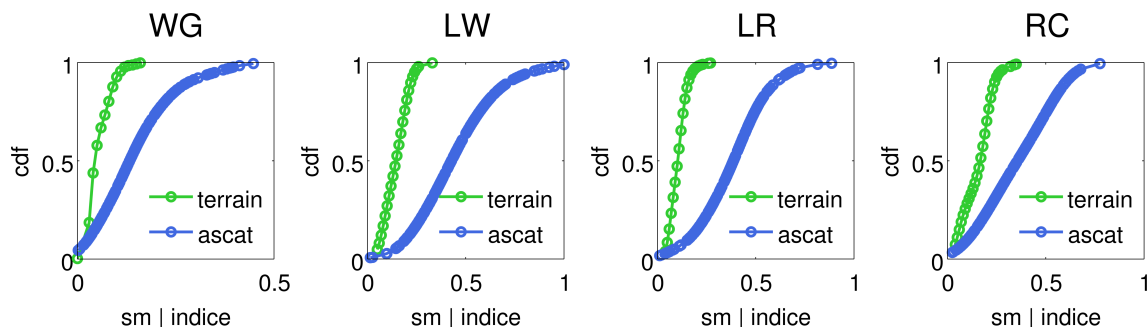


FIGURE 3.14 – Distributions cumulées des jeux de données ASCAT et des mesures terrain sur les quatre bassins versants pour 2010.

	AM		PM	
	R	IC	R	IC
WG	0.58	0.35-0.74	0.45	0.22-0.63
LW	0.72	0.57-0.83	0.60	0.42-0.73
LR	0.22	*	0.35	0.08-0.57
RC	0.70	0.53-0.82	0.53	0.34-0.68

Tableau 3.4 – Résultats statistiques (corrélations R et intervalles de confiance IC) pour les données ASCAT normalisées statistiquement avec les mesures terrain de 2010 (* : non significatif statistiquement, valeur $p \geq 0.05$).

dans le but de respecter la distribution des valeurs d'humidité dans une certaine gamme.

Dans un premier temps, les distributions statistiques sont calculées pour le jeu de données ASCAT et pour les mesures terrain sur chaque bassin versant (fig. 3.14). Une distribution cumulée, ou *cumulative density function* (cdf) en anglais, représente le pourcentage de données qui sont inférieures à une valeur. Prenons l'exemple de Little Washita (LW) sur la figure 3.14, on lit que 50% des données ASCAT ont une valeur inférieure à 0.5 alors que 50% des mesures terrain ont une valeur inférieure à 0.2 m³/m³. On remarque d'ailleurs la différence de distribution entre ASCAT et les mesures terrain. Faire correspondre ces distributions revient à avoir les mêmes pourcentages de données qui sont inférieurs aux mêmes valeurs seuils (fig. 3.14).

Une fois cette transformation réalisée, les nouvelles séries temporelles des indices d'humidité d'ASCAT peuvent être comparées aux mesures terrain (fig 3.15) et les coefficients de corrélation ont été calculés de la même manière que précédemment ainsi que les intervalles de confiance associés (tab. 3.4). Ces nouveaux coefficients sont très proches voire identiques à ceux calculés lors de la normalisation linéaire indiquant que la normalisation statistique n'apporte aucun élément supplémentaire à l'analyse déjà menée.

3.3.6 Dépendance d'AMSR-E à la saisonnalité

Les biais indiqués dans le tableau 3.2 et la figure 3.13 sont des valeurs moyennes sur l'année 2010. Les biais des produits SMOS et ECMWF sont constants tout au long de l'année mais le biais du produit VUA varie avec les saisons. Sur trois des quatre bassins versants, la tendance est identique : un biais très réduit en été et un biais conséquent en hiver (figs. 3.8, 3.9, 3.10 et 3.11).

Le produit d'humidité VUA est retrouvé à partir des températures de brillance d'AMSR-E à 6.9 GHz et 10.7 GHz (l'une ou l'autre fréquence est utilisée en fonction des RFI présents ou non). A ces fréquences, les observations sont beaucoup plus sensibles à la végétation que

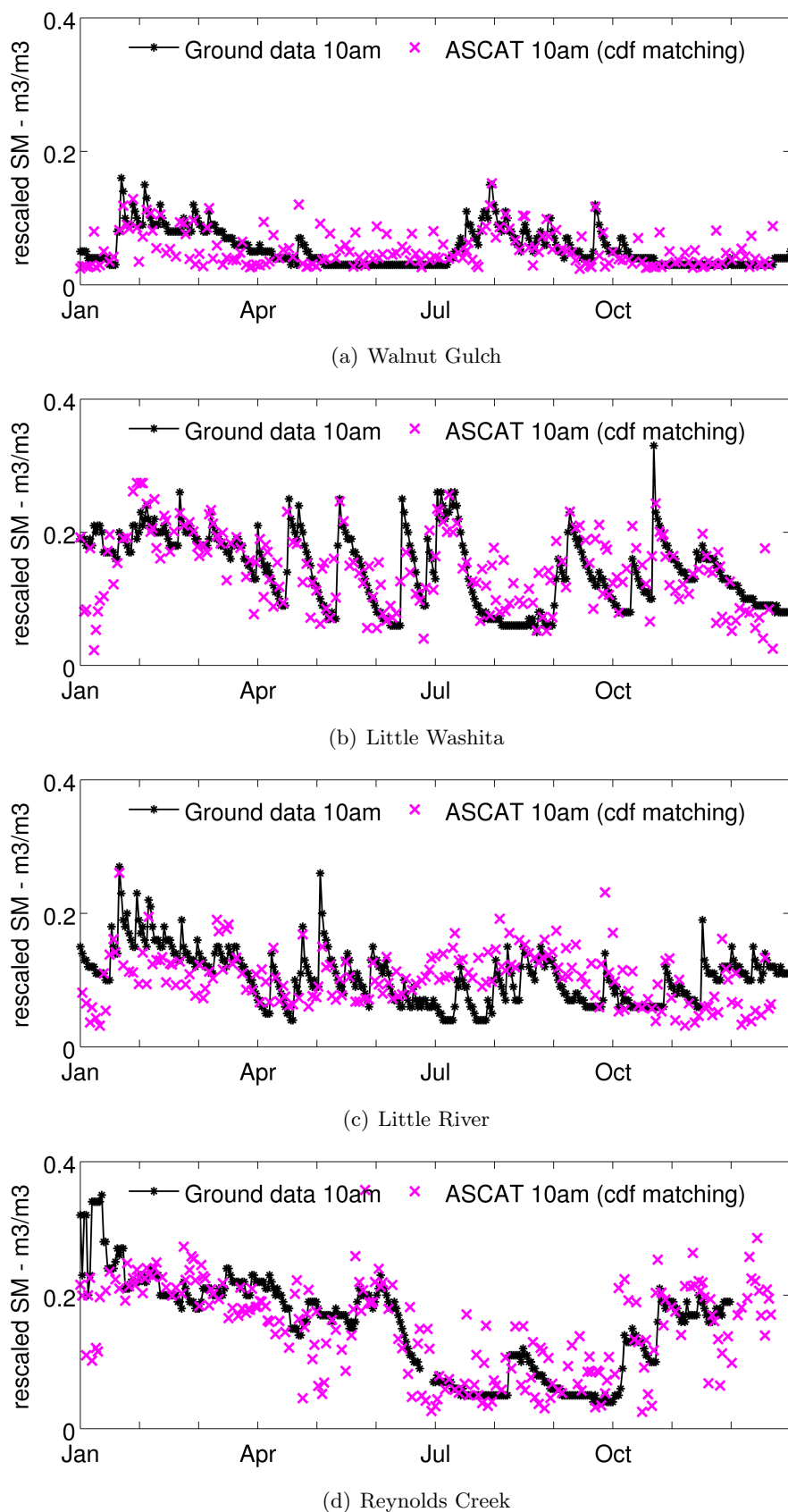


FIGURE 3.15 – Série temporelle de l'humidité du sol d'ASCAT normalisée statistiquement avec les mesures terrain sur les quatre bassins versants en 2010.

la bande L de SMOS (fig. 2.3, p.16). A cause de cette grande sensibilité à la végétation, une comparaison directe entre le biais de VUA et le LAI est montrée en fig. 3.16. Comme toute température de brillance, les acquisitions d'AMSR-E sont également sensibles à la température de surface, donc ce paramètre physique a également été comparé au biais sur la fig. 3.16.

Le produit LAI d'ECOCLIMAP a été utilisé et n'est disponible que tous les dix jours alors que la température du sol a été mesurée chaque jour par les stations terrain. Quand la corrélation est statistiquement significative, une régression linéaire est réalisée et un intervalle de confiance à 95% sur le coefficient de corrélation a été calculé. Ces intervalles sont naturellement assez larges pour la comparaison avec le LAI à cause du faible nombre de points pour 2010 (entre 23 et 32).

D'après la figure 3.16, le biais du produit VUA diminue lorsque le LAI et la température du sol augmentent. Le LAI et la température du sol peuvent être deux paramètres physiquement très corrélés donc il est naturel de retrouver la même tendance quand on les compare à une même variable. La dépendance négative au LAI est un résultat surprenant car si les fréquences d'AMSR-E utilisées sont plus sensibles à la végétation, l'humidité du sol devrait être correctement retrouvée lorsqu'il n'y a pas ou très peu de végétation. Or, cette étude montre le contraire. Cela signifierait alors que l'algorithme de VUA ne corrige pas correctement l'effet de la végétation. En début d'été lorsque la végétation est la plus développée, le biais est très bas montrant ainsi que la végétation a bien été corrigée. Au contraire, en hiver lorsqu'il n'y a plus beaucoup de végétation, le biais est bien plus élevé (fig. 3.9 par exemple). Dans l'algorithme de VUA, l'épaisseur optique de la végétation, qui représente la quantité de signal absorbé par la végétation, est calculée à partir du rapport de polarisation (différence entre les polarisations verticale et horizontale divisée par leur somme, Antoon et al., 2005). On peut supposer qu'une seule observation à un seul angle d'incidence ne contient pas assez d'information pour retrouver à la fois l'humidité et l'épaisseur optique de la végétation.

La température du sol est calculée à partir de la température de brillance acquise à 36 GHz en polarisation verticale à travers une relation linéaire empirique (Holmes et al., 2009). L'erreur standard d'estimation de modèle empirique a été trouvée autour de 3.5 K pour une surface à faible végétation (Holmes et al., 2009). Cependant, s'il y avait un biais constant dans la température du sol calculée, cela n'impacterait pas la dépendance saisonnière du biais. En modifiant le niveau moyen de la température du sol, cela ne fait que modifier le niveau moyen de l'humidité retrouvée mais ne modifie en rien l'évolution du biais au fil des saisons.

3.3.7 Sensibilité de VUA

Afin de déterminer plus précisément la dépendance saisonnière du biais du produit VUA, une étude de sensibilité portant sur un ensemble de températures de brillance a été menée. En effet, l'algorithme VUA ou LPRM prend en entrée les températures de brillance d'AMSR-E à 6.9 et 10.7 GHz pour retrouver l'humidité du sol ainsi que la fréquence 36.9 GHz pour calculer la température efficace du sol qui est utilisée pour calculer cette humidité.

Pour cette étude, les températures de brillance à 6 GHz et 36 GHz d'AMSR-E ont été extraites afin de retrouver l'humidité, l'épaisseur optique de la végétation et la température du sol avec l'algorithme de VUA. Le centre du bassin versant de Little Washita a été choisi pour cet exemple. Les séries temporelles des températures de brillance, du rapport de polarisation sont montrées sur la figure 3.17 et les paramètres retrouvés par l'algorithme VUA sont comparés aux données terrain d'humidité et de température et au LAI ECOCLIMAP. Les données terrain représentent des moyennes de stations et elles sont comparées à un seul pixel d'AMSR-E, ce qui peut représenter une source d'erreur mais le but de cette étude est avant tout d'identifier l'influence de la végétation et de la température du sol sur les valeurs des paramètres retrouvés de VUA.

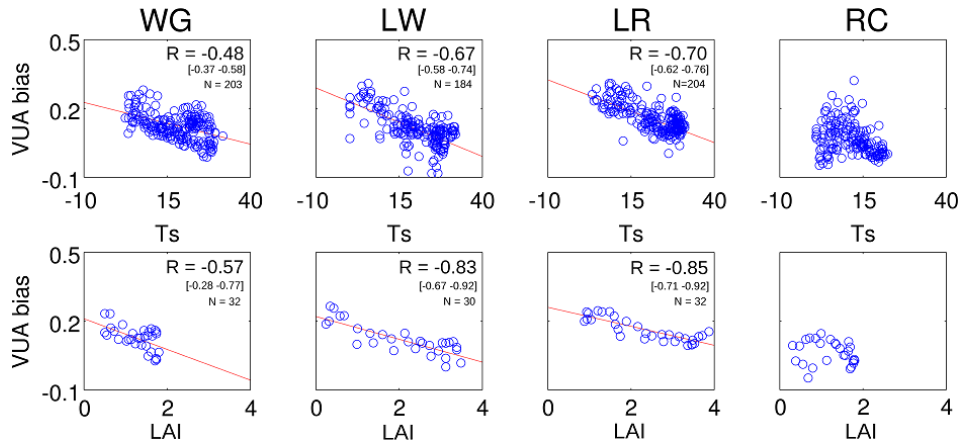


FIGURE 3.16 – Comparaison entre le biais du produit VUA et la végétation (LAI) et la température de surface (Ts). Le biais est anti-corrélé au LAI et à la température du sol. Aucune régression n'a été réalisée sur Reynolds Creek car elle n'était pas statistiquement significative. R est le coefficient de corrélation, [.] est son intervalle de confiance à 95% et N est le nombre de points.

En comparant les humidités retrouvées par VUA et les données terrain, on retrouve ce biais qui varie au fil des saisons : plus important en hiver qu'en été alors que la végétation est la plus dense (exemple en fig. 3.10). Les comparaisons concernant la végétation et la température du sol soulèvent plusieurs interrogations :

- l'épaisseur optique retrouvée par l'algorithme VUA ne comporte aucun signe de saisonnalité alors que l'indice LAI d'ECOCLIMAP en indique clairement un. Ce paramètre est retrouvé à l'aide du rapport de polarisation à 6 GHz dans l'algorithme et ce rapport n'indique pas de saisonnalité non plus. L'épaisseur optique de la végétation peut elle réellement être retrouvée en utilisant le rapport de polarisation à cette fréquence ?
- la température du sol est clairement sur-estimée d'un peu plus de 10 K. En revanche, cette sur-estimation est constante tout au long de l'année et la dynamique de la température du sol est très bien retranscrite ici. Est ce qu'un biais constant dans la modélisation de la température du sol peut influencer sur le biais saisonnal de l'humidité ?

Pour répondre à ces questions, deux tests ont été réalisés en divisant par 2 l'épaisseur optique et en appliquant un biais à la température du sol modélisée pour identifier les effets sur l'humidité finale retrouvée par l'algorithme VUA. Initialement, 10K devaient être retranchés à la température du sol mais plus aucune valeur d'humidité n'était retrouvée. Souhaitant voir l'effet d'un biais constant, le test a été mené en ajoutant un biais de 10K à la température du sol.

Le premier test sur la végétation consiste à diviser par 2 l'épaisseur optique de la végétation après qu'elle ait été calculée avec le rapport de polarisation. En diminuant donc l'effet de la végétation, on remarque une diminution générale du niveau de l'humidité retrouvée sur la figure 3.18(a). Originellement, le biais était plus fort en hiver qu'en été. Cela voudrait dire alors que l'effet de la végétation est sur-estimée pendant l'hiver et qu'il faudrait diminuer l'épaisseur optique de la végétation.

Le second test concerne la modélisation de la température du sol où 10K ont été ajoutés à la température modélisée avec la fréquence 36 GHz. En augmentant la température du sol, on constate une augmentation générale de l'humidité du sol (fig. 3.18(b)). Remarquant que le biais entre la température modélisée et la température mesurée est stable tout au long de l'année, modifier le niveau général de la température n'a comme incidence qu'une modification du niveau

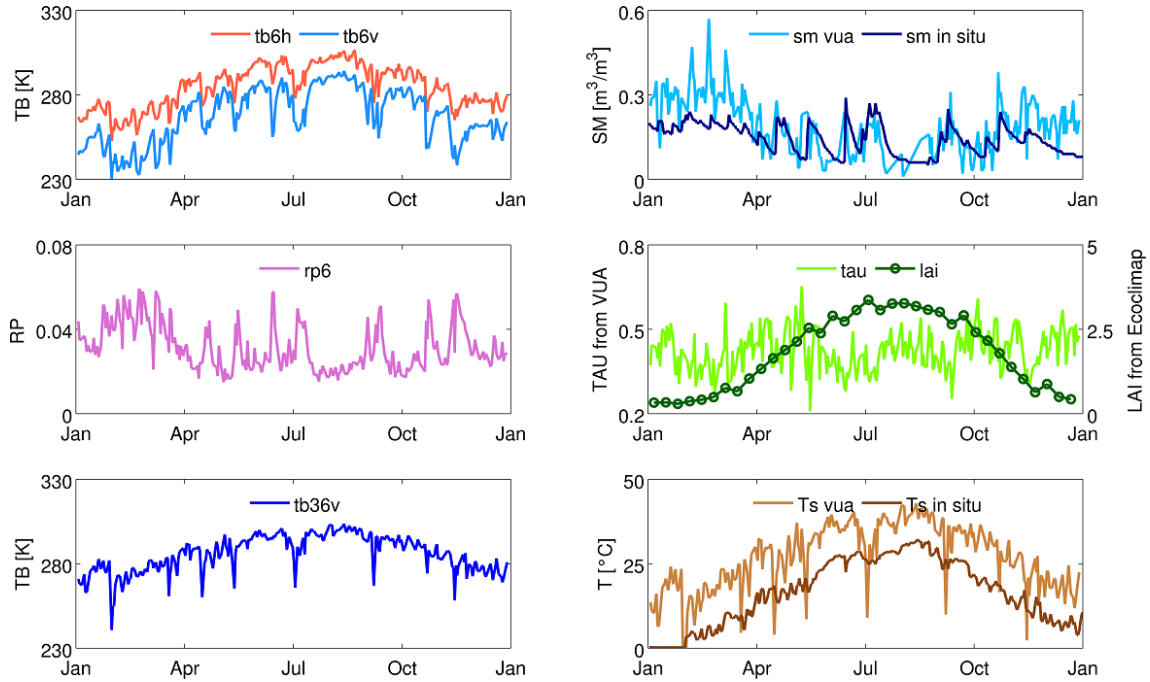


FIGURE 3.17 – Températures de brillance AMSR-E (Tb 6 GHz et 36 GHz), rapport de polarisation (RP) à 6 GHz (différence des polarisations divisée par leur somme), humidité des sols (sm) retrouvée par VUA et mesurée sur le site de Little Washita, épaisseur optique de la végétation (tau) retrouvée par VUA et LAI d'ECOCLIMAP, température du sol (T) retrouvée par VUA et mesurée sur le site.

général de l'humidité et ne résout pas le problème du biais saisonnier.

La piste d'un problème de modélisation de la végétation est donc l'axe privilégié pour une future amélioration de l'algorithme VUA. Il est fort possible que le rapport de polarisation d'un seul angle ne contienne pas assez d'information pour retrouver les valeurs d'humidité et de végétation en même temps. Le chapitre 6 tentera de répondre à ce problème en adaptant l'algorithme VUA aux observations multi-angulaires de SMOS.

3.4 Aperçu des produits de végétation SMOS sur les bassins versants tests

L'algorithme de SMOS retrouve l'humidité à partir des observations multi-angulaires du radiomètre mais également l'épaisseur optique de la végétation qui représente la quantité de signal absorbée par la couche de végétation en raison de son contenu en eau. L'humidité et l'épaisseur optique sont optimisées en même temps et même si l'algorithme permet de ne retrouver que l'un ou l'autre, dans la plupart des cas ces deux variables sont retrouvées conjointement. Une densification de la végétation résulte en une augmentation de l'épaisseur optique qui a pour conséquence de diminuer l'écart entre les polarisations horizontales et verticales et rend plus complexe le fait de retrouver l'humidité.

Des séries temporelles de l'épaisseur optique de la végétation TAU peuvent ainsi être tracées de la même manière que pour l'humidité des sols sur les quatre bassins versants tests. Il a déjà été montré lors de plusieurs études (Jackson et al. (2012)) que cette variable est très bruitée et qu'elle n'est pas utilisable pour des études sur la végétation telle quelle. Pour cela, Hornbuckle (2012) propose d'effectuer un lissage temporel qui consiste à moyennner TAU sur une fenêtre

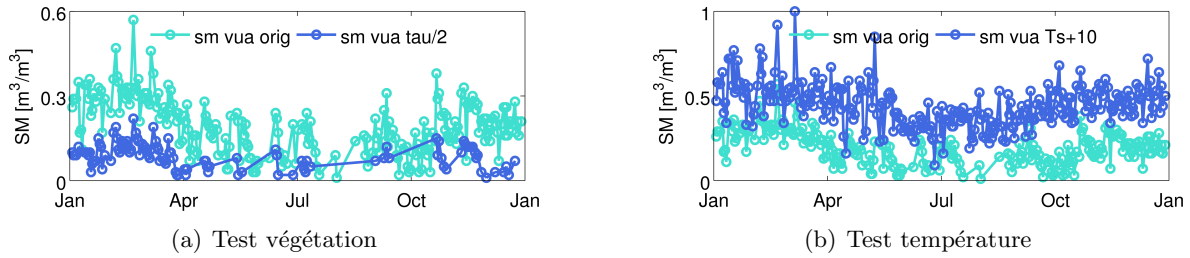
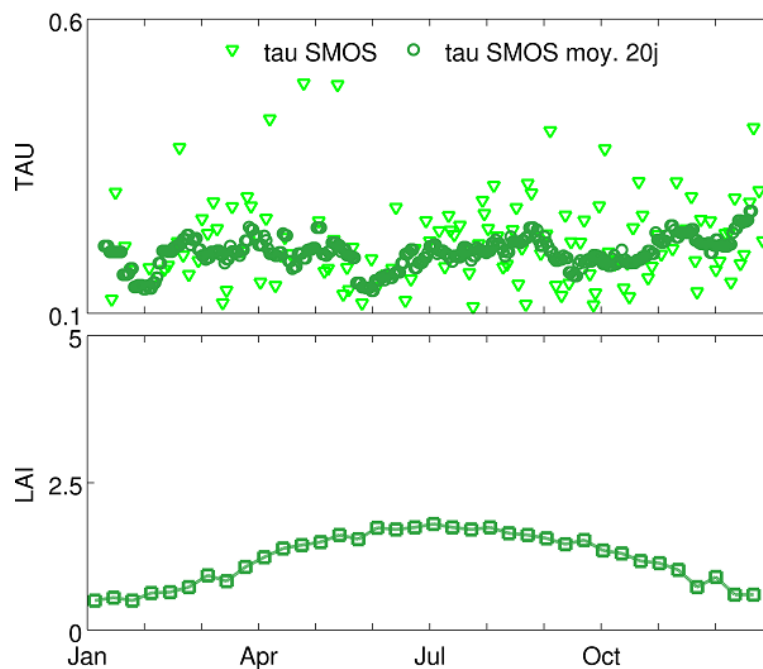


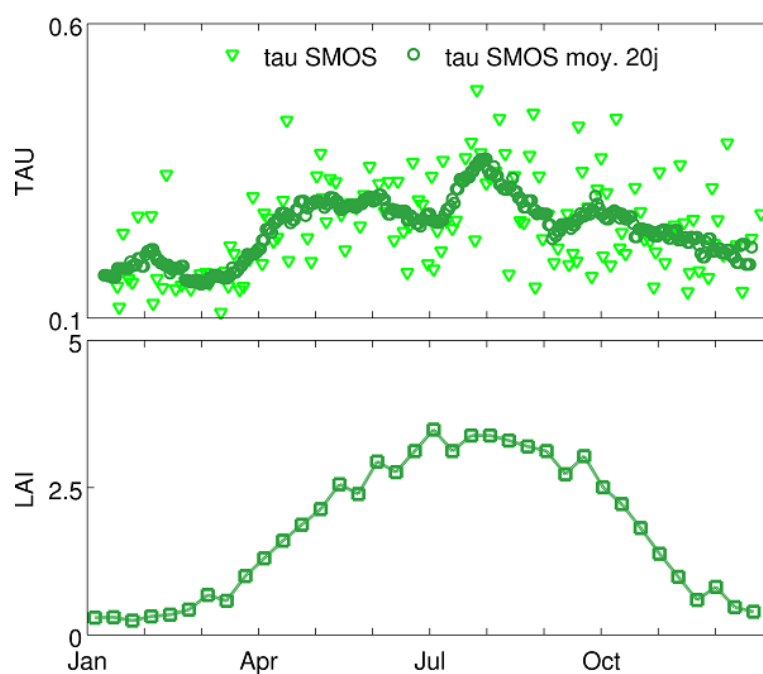
FIGURE 3.18 – Tests de sensibilité de l'algorithme VUA sur l'effet de la végétation (a) et de la température du sol (b).

glissante de quelques jours afin de réduire le bruit de cette variable. Les séries temporelles de TAU et du TAU lissé sur 20 jours sont présentées sur les figures 3.19, et 3.20. Les indices de végétation LAI d'ECOCCLIMAP ont également été ajoutés afin d'avoir une référence et une comparaison n'ayant pas de données terrain concernant la couverture de la végétation sur les sites tests.

En lissant l'épaisseur optique de la végétation sur 20 jours, une dynamique semble se dessiner, notamment sur Little Washita (fig. 3.19(b)). Ces séries temporelles de la végétation ne sont pas utilisables pour le moment car elles sont encore trop bruitées mais un premier lissage laisse apparaître des tendances qui sont encourageantes. Le niveau 2 de SMOS optimise à la fois l'humidité des sols et l'épaisseur optique de la végétation et très souvent, le bruit de mesure lors des acquisitions des températures de brillance sont reportées sur la seconde variable optimisée, i.e. l'épaisseur optique. Ces résultats mettent en avant un réel besoin de mieux modéliser cette couche de végétation qui impacte également la valeur d'humidité retrouvée.

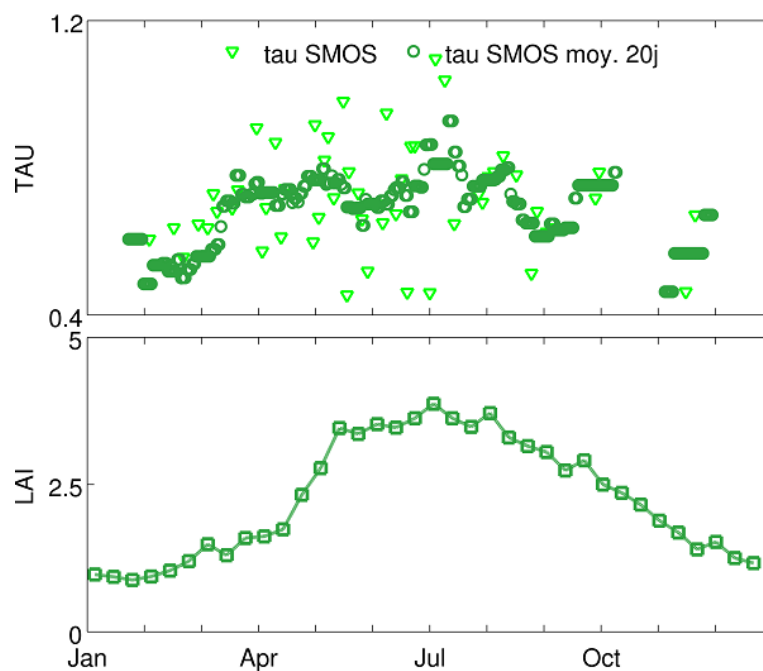


(a) Walnut Gulch

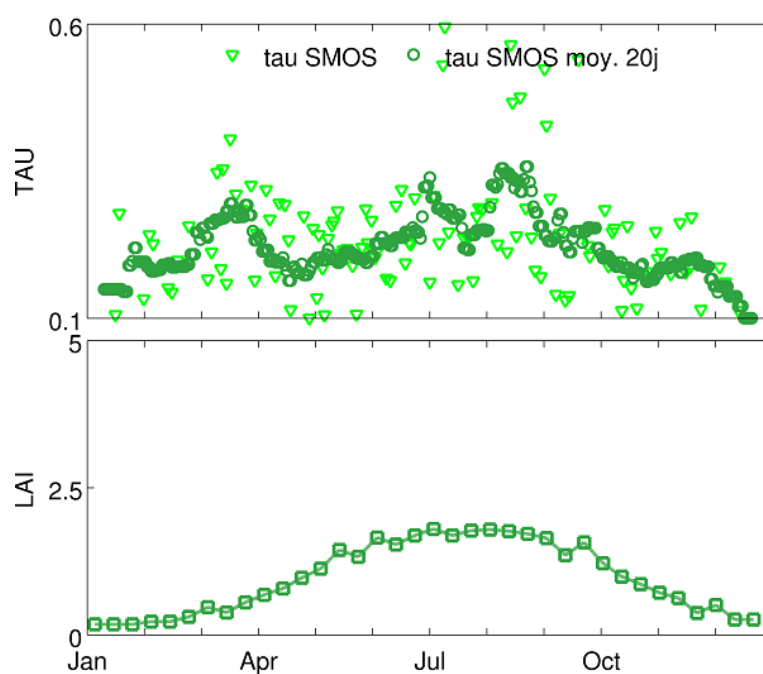


(b) Little Washita

FIGURE 3.19 – Série temporelle en 2010 de l'épaisseur optique de la végétation retrouvée par SMOS puis lissée sur 20 jours et accompagnée du LAI d'ECOCLIMAP afin de comparer l'évolution de la végétation sur Walnut Gulch et Little Washita.



(a) Little River



(b) Reynolds Creek

FIGURE 3.20 – Série temporelle en 2010 de l'épaisseur optique de la végétation retrouvée par SMOS puis lissée sur 20 jours et accompagnée du LAI d'ECOCLIMAP afin de comparer l'évolution de la végétation sur Little River et Reynolds Creek.

3.5 Conclusions et perspectives

Cette étude a été menée dans le cadre des travaux de validation des produits d'humidité SMOS. Les données SMOS ont été comparées à des mesures terrain et à d'autres produits d'humidité retrouvée à partir d'observations satellite et de modèle. Les données SMOS, VUA, ECMWF et ASCAT ont été inter-comparées sur quatre bassins versants aux Etats-Unis couvrant une large variété de climats et de caractéristiques du sol.

Sur les quatre sites test, aucun produit n'a obtenu une corrélation significativement meilleure que les autres. Puisque cette étude ne couvrait que l'année 2010, seulement 44 à 72 dates communes à tous les jeux de données ont été utilisées pour calculer les statistiques, ce qui s'est traduit par de larges intervalles de confiance et par une quasi-équivalence de toutes les valeurs de corrélation calculées. Les mesures statistiques utilisées pour différencier les quatre jeux de données ont été l'erreur standard d'estimation (SEE), le biais et l'erreur globale (RMSE). La relation entre ces trois mesures est rappelée : $SEE^2 + \text{biais}^2 = RMSE^2$.

Sur la majorité des sites test, ECMWF a obtenu une SEE significativement plus basse ($0.038 \text{ m}^3/\text{m}^3$ en moyenne) que SMOS et VUA ($0.052 \text{ m}^3/\text{m}^3$ et $0.061 \text{ m}^3/\text{m}^3$ en moyenne), ce qui représente un ensemble de données moins dispersé que les autres autour des mesures terrain. La dynamique de l'humidité ECMWF était très proche de celle mesurée par les stations installées sur les bassins versants avec des événements pluvieux clairement identifiables ainsi que les périodes de drainage. Cependant, SMOS a nettement obtenu le biais le plus faible avec une moyenne de $0.029 \text{ m}^3/\text{m}^3$ alors que ECMWF a obtenu un biais moyen de $0.150 \text{ m}^3/\text{m}^3$ et VUA de $0.115 \text{ m}^3/\text{m}^3$. SMOS a également obtenu l'erreur globale la plus faible sur tous les sites sauf Reynolds Creek. L'erreur globale de SMOS était de $0.062 \text{ m}^3/\text{m}^3$ contre $0.156 \text{ m}^3/\text{m}^3$ pour AMSR-E et $0.130 \text{ m}^3/\text{m}^3$ pour VUA. Bien que l'objectif de la mission SMOS, qui est de retrouver l'humidité avec une précision de $0.04 \text{ m}^3/\text{m}^3$, ne soit pas atteint sur ces quatre bassins versants en 2010, SMOS a amélioré l'erreur totale d'un facteur 2 au minimum en moyenne pour une corrélation équivalente aux autres produits d'humidité.

Pour tous les sites, excepté Little River l'après-midi, l'erreur standard d'estimation de SMOS était plus importante que le biais alors que c'était l'opposé pour les autres produits d'humidité. Cela signifie que les différences dans la variabilité des séries temporelles de SMOS avec les mesures terrain sont plus importantes que le biais à long terme. L'assertion contraire est vraie pour les produits ECMWF et VUA.

L'analyse des séries temporelles a révélé un biais constant dans le jeu de données ECMWF : $0.17 / 0.10 / 0.26 / 0.07 \text{ m}^3/\text{m}^3$ pour les bassins versants WG / LW / LR / RC. Même après la mi-Novembre 2010, quand le modèle de surface fut modifié, ce biais était encore présent sur tous les sites mais avait légèrement diminué sur Walnut Gulch. Malgré de très bons coefficients de corrélation, ce niveau minimum d'humidité est un obstacle majeur à l'utilisation des données ECMWF en absolu sur ces quatre sites. Le produit VUA a présenté également un biais mais qui n'était pas constant et paraissait évoluer avec les saisons. Le biais était plus grand en hiver quand la végétation était la moins développée et quand la température était à son minimum et le biais était quasi nul quand la végétation était la plus développée et quand la température était à son maximum. L'évolution saisonnière est supposée être liée à la végétation à cause des fréquences utilisées d'AMSR-E dans l'algorithme VUA (6.9 GHz et 10.7 GHz) qui sont plus sensibles à la couche de végétation que SMOS à 1.4 GHz. Cependant il n'est pas exclu que ce biais provienne d'une modélisation non précise de la température du sol dans VUA qui est effectuée en utilisant la fréquence 36 GHz d'AMSR-E. Les observations mono-angulaires d'AMSR-E ne comportent peut être pas assez d'information pour retrouver à la fois l'humidité du sol et l'épaisseur optique de la végétation. L'indice d'humidité ASCAT était très bruité comparé aux autres produits et en dépit de la bonne observation de quelques événements pluvieux, la dynamique et l'évolution

globale de l'humidité n'étaient pas bien représentées par ASCAT sur les quatre bassins versants.

La comparaison des performances statistiques des différents produits entre le matin et l'après-midi ont révélé que le modèle ECMWF ne présentait aucune différence alors que les produits satellite SMOS et ASCAT ont montré des différences importantes : $R = 0.80 / 0.65$, $SEE = 0.043 / 0.061 \text{ m}^3/\text{m}^3$, $\text{biais} = 0.027 / 0.030 \text{ m}^3/\text{m}^3$, $\text{RMSE} = 0.052 / 0.073 \text{ m}^3/\text{m}^3$ pour SMOS et $R = 0.67 / 0.49$ pour ASCAT (matin / après-midi en moyenne sur les quatre sites). Ces différences de performance peuvent être expliquées par le quasi-équilibre thermique du sol le matin qui rend tout à fait comparable des observations satellites des 2-3 premiers centimètres du sol avec des mesures terrain à 5 cm. En revanche l'après-midi, le sol n'est plus en équilibre et comparer des observations satellite avec des mesures terrain à des profondeurs différentes n'est plus aussi correct que pour le matin. Une autre explication porterait sur la modélisation des températures du sol et de la végétation dans les algorithmes. Elles sont supposées égales et cette hypothèse n'est plus vraie l'après-midi où la température de la végétation est attendue plus proche de la température de l'air que de la température du sol. Concernant les différences de performance du produit SMOS, une dernière explication possible est la présence d'interférences. Les probabilités d'occurrence ont été plus élevées l'après-midi en 2010 sur tous les sites sauf Reynolds Creek.

La normalisation des données ASCAT a été effectuée dans un premier temps linéairement puis de manière statistique en faisant correspondre sa distribution statistique avec celle des mesures terrain. Statistiquement, les normalisations linéaires et statistiques ont obtenu des résultats équivalents. Malgré ces normalisations, les données ASCAT restaient plus bruitées.

Les résultats de cette étude ont abouti à un article qui a été soumis pour la première fois en Août 2011 au journal *Remote Sensing of Environment* se nommant *Comparison between SMOS and other satellite and model forecast soil moisture products and validation over four watersheds in the US*. Les auteurs de cet article sont Delphine J. Leroux, Yann H. Kerr, A. Al Bitar, Claire Gruhier (CESBIO, Toulouse), Rajat Bindlish, Thomas J. Jackson (USDA, Beltsville, Etats-Unis), Béatrice Berthelot (Magellium, Toulouse), Gauthier Portet (Telespazio, Toulouse).

J'ai également participé à plusieurs autres travaux de validation des données SMOS dont les conclusions sont semblables à celles présentées dans cette étude :

- *Evaluation of SMOS soil moisture products over continental US using the SCAN/SNOTEL network* accepté dans le journal *IEEE Transactions on Geoscience and Remote Sensing* et écrit par Al Bitar A., Leroux D.J., Kerr Y.H., Merlin O., Richaume P., Sahoo A. et Wood E.F.
- *Validation of Soil Moisture and Ocean Salinity (SMOS) soil moisture over watershed networks in the U.S.* accepté dans le journal *IEEE Transactions on Geoscience and Remote Sensing* et écrit par Jackson T.J., Bindlish R., Cosh M., Zhao T., Starks P., Bosch D., Seyfried M., Moran S., Kerr Y.H. et Leroux D.J.
- *An initial assessment of SMOS derived soil moisture over the continental United States* accepté dans le journal *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* et écrit par Pan M., Sahoo A.K., Wood E.F., Al Bitar A., Leroux D.J. et Kerr Y.H.
- *The SMOS soil moisture retrieval algorithm* accepté dans le journal *IEEE Transactions on Geoscience and Remote Sensing* et écrit par Kerr Y.H., Waldteufel P., Richaume P., Mahmoodi A., Wigneron J.-P., Ferrazzoli P., Al Bitar A., Cabot F., Leroux D.J., Mialon A. et Delwart S.

Chapitre 4

Validation des données SMOS à l'échelle globale et sources possibles des erreurs SMOS

La mission SMOS (*Soil Moisture and Ocean Salinity*) produit des données depuis maintenant plus de deux ans et dans le cadre du processus de validation de ces données, comparer ce nouveau jeu de données d'humidité aux produits déjà existants est maintenant possible à l'échelle globale. Dans ce chapitre, le produit d'humidité SMOS (v400) est évalué à l'échelle globale en utilisant la méthode de la triple collocation pour l'année 2010. Cette méthode statistique est basée sur la comparaison de trois jeux de données et produit des cartes globales d'erreur en inter-comparant statistiquement les variations temporelles de ces trois jeux de données. Cette méthode a été appliquée aux produits d'humidité suivants : SMOS, AMSR-E/VUA (*Vrije University of Amsterdam*), AMSR-E/NSIDC (*National Snow and Ice Data Center*), ASCAT (*Advanced Scatterometer*) et ECMWF (*European Center for Medium range Weather Forecasting*). Les erreurs produites par la triple collocation ne sont pas des erreurs absolues mais relatives au choix des trois jeux de données. Cependant, cette étude montre que la structure spatiale des erreurs SMOS sont indépendantes de ce choix et soulignent les mêmes régions du globe où les erreurs SMOS sont les plus faibles et les plus élevées.

Cette carte globale de l'erreur SMOS est ensuite reliée à différents paramètres biophysiques comme la texture du sol (quantités de sable et d'argile), la couverture du sol (forêt, lacs,...) ainsi que la probabilité moyenne d'occurrence des interférences en 2010. A l'échelle globale, la fraction de forêt vue par le radiomètre a la plus grande influence sur l'erreur SMOS (près de 57%) alors que les interférences ne représentaient finalement que 1.7% selon l'analyse de la variance (ANOVA) effectuée à partir d'un modèle de régression linéaire multiple. Ces pourcentages d'influence évoluent significativement suivant les continents et de nombreuses différences sont à noter : la texture du sol a la plus grande influence en Europe alors que ce sont les interférences sur le continent asiatique. Enfin une classification a permis d'isoler des ensembles de paramètres avec des valeurs seuils qui mènent aux erreurs les plus fortes et les plus faibles.



4.1 Introduction

4.1.1 Quelle méthode pour une étude globale ?

De nombreuses études ont été réalisées ou sont en cours pour la validation des données SMOS (Al Bitar et al., 2012, Bircher et al., 2012, Gherboudj et al., 2012, Parrens et al., 2011, Peischl et al., 2009, Schlenz et al., 2012). La plupart porte sur une comparaison des données d'humidité SMOS avec des mesures réalisées sur le terrain ou avec des observations réalisées lors de campagnes aéroportées.

Al Bitar et al. (2012) a utilisé un réseau de stations situé en Amérique du Nord appelé SCAN (*Soil Climate Analysis Network*) et le réseau SNOTEL (*Snowpack Telemetry*) pour comparer les données d'humidité SMOS avec les mesures réalisées par ces stations météorologiques. Cette étude portait sur 238 stations et montre que le produit SMOS retrouve très bien la dynamique des mesures terrain mais qu'il a tendance à sous-estimer l'humidité.

Jackson et al. (2012) a étudié l'humidité et l'épaisseur optique de la végétation de SMOS sur les quatre bassins versants aux Etats-Unis décrits dans le chapitre précédent. Ils ont conclu que SMOS avait presque réussi son objectif de précision concernant l'humidité avec une erreur globale sur la comparaison du matin de $0.043 \text{ m}^3/\text{m}^3$ et de $0.047 \text{ m}^3/\text{m}^3$ pour l'après-midi alors que l'épaisseur optique de la végétation n'est pas bien retrouvée et est trop bruitée pour pouvoir être utilisée dans une analyse de la végétation.

Il existe bien d'autres études de validation des données SMOS encore en cours notamment sur un bassin versant du Danube en Allemagne (Schlenz et al., 2012), au Danemark (Bircher et al., 2012), en Chine, au Tibet, en Australie (Peischl et al., 2009) et en France (Parrens et al., 2011).

Toutes ces études s'appuient donc sur une comparaison des données SMOS avec une ou plusieurs stations. Mais aussi nombreuses que sont ces stations, on ne peut généralement pas conclure pour la globalité du produit SMOS car chaque site où ont été effectuées les mesures, est très spécifique avec un type de sol, une occupation du sol et un climat qui lui sont propres. De plus il est toujours difficile de comparer une observation satellite qui a une résolution de plusieurs kilomètres avec une mesure ponctuelle d'une station.

Pour évaluer un jeu de données à l'échelle globale, il faut alors le comparer à un jeu de données similaire. Comparer SMOS et AMSR-E par exemple, permettrait d'identifier leurs différences, leur biais qui pourraient éventuellement évoluer dans le temps, mais il serait impossible de conclure sur la qualité de ces produits. Pour cela, un troisième jeu de données est introduit. On peut alors supposer que si deux des trois produits constatent une augmentation de l'humidité des sols et pas le troisième alors ce dernier sera considéré comme moins fiable car il n'est pas en accord avec la majorité.

Pour quantifier et spatialiser cette fiabilité, une méthode statistique est nécessaire pour réaliser cette étude à l'échelle globale. La méthode utilisée est la triple collocation (Stoffelen, 1998) et permet de comparer trois jeux de données afin d'identifier leur biais et leurs erreurs sur l'ensemble du globe. La triple collocation est une méthode basée sur les anomalies temporelles et on parlera donc d'erreur variable. La partie *biais* de l'erreur n'est pas prise en compte par cette méthode.

4.1.2 Objectifs

Les deux objectifs de cette étude sont d'évaluer la précision relative du produit d'humidité de SMOS comparé à d'autres jeux de données globaux d'humidité pour en déduire une spatialisation des erreurs relatives de SMOS et ensuite de relier cette erreur à des paramètres physiques.

Pour cela, la méthode de triple collocation est appliquée aux produits satellite SMOS,

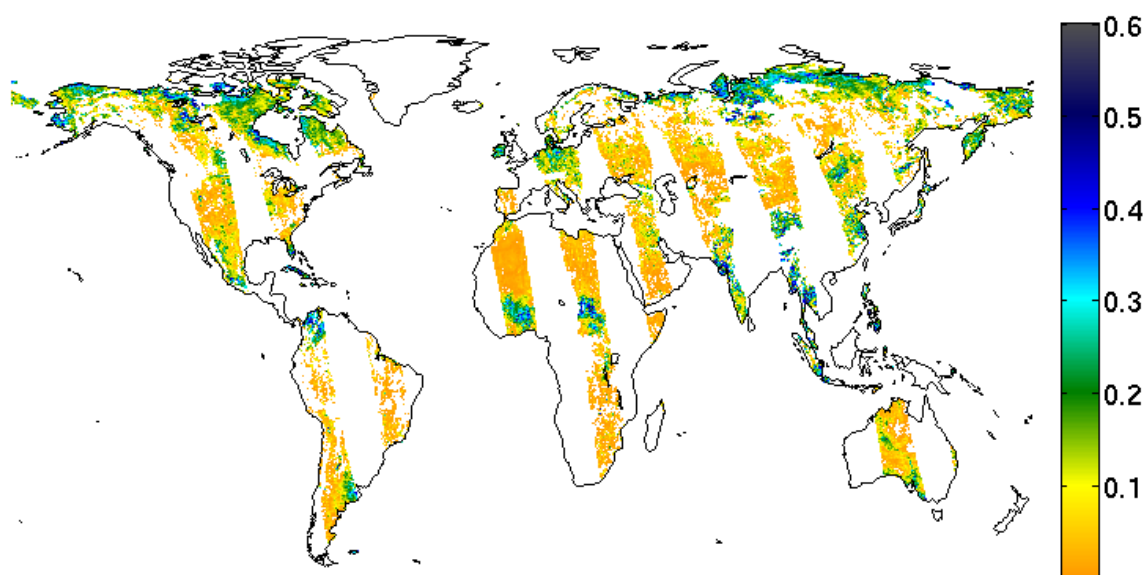


FIGURE 4.1 – Humidité des sols (m^3/m^3) retrouvée par SMOS le 15 Septembre 2010 à 6:00h, heure solaire locale.

AMSR-E (VUA et NSIDC), ASCAT et au produit du modèle ECMWF pour l'année 2010. Dans un second temps, une analyse de la variance permettra d'identifier les possibles sources des erreurs SMOS à l'échelle globale puis continentale.

4.1.3 Jeux de données utilisés

Les données ont été présentées en détails dans le chapitre précédent et uniquement un bref résumé des produits utilisés est donné dans cette partie.

SMOS observe la Terre à divers angles d'incidence, de 0° à 55° et c'est cette évolution angulaire des températures de brillance acquises qui permet de retrouver l'humidité du sol et l'épaisseur optique de la végétation. Avec son radiomètre en bande L, le signal SMOS représente les 2 à 5 premiers centimètres du sol. Ces données sont disponibles sur la grille ISEA-4h9 dont les nœuds sont tous espacés de 15 km. Dans cette partie, la version 4 de ces produits est utilisée.

Deux produits dérivés des observations d'AMSR-E sont testés dans cette étude : le produit NSIDC et le produit VUA. Le premier produit est celui distribué par le NSIDC (*National Snow and Ice Data Center*) et est obtenu à partir des fréquences 10.7 GHz et 18.7 GHz (Njoku et al., 2003). Cet algorithme était initialement prévu pour la fréquence 6.9 GHz mais à cause d'une trop grande quantité d'interférence, cette fréquence n'a pas pu être utilisée. L'humidité du sol et l'épaisseur optique de la végétation sont distribuées sur une grille régulière de 25 km (Njoku, 2004). Dans cette partie, le produit Level 3 DailyLand V6 est utilisé.

Le produit VUA (Vrije University of Amsterdam) est basé sur un modèle de transfert radiatif qui utilise les bandes C et X d'AMSR-E (la bande X est seulement utilisée lorsque la bande C est trop contaminée par les interférences) et qui permet de retrouver l'humidité et l'épaisseur optique de la végétation ainsi que la température du sol grâce à l'utilisation de la fréquence 36.5 GHz d'AMSR-E. Le produit Level 3 AMSR-E v03 Grid est utilisé et est disponible sur une grille de 0.25° . En bande C, le signal provient essentiellement des 2 premiers centimètres du sol alors qu'en bande X on estime que seul l'état de surface est observé.

ASCAT est le seul produit de cet étude à être basé sur des observations actives en bande C. Un indice d'humidité entre 0 et 1 est retrouvé en utilisant l'algorithme de l'Université de Vienne (Wagner et al., 1999, Naeimi et al., 2009) qui permet de corriger l'effet de la végétation

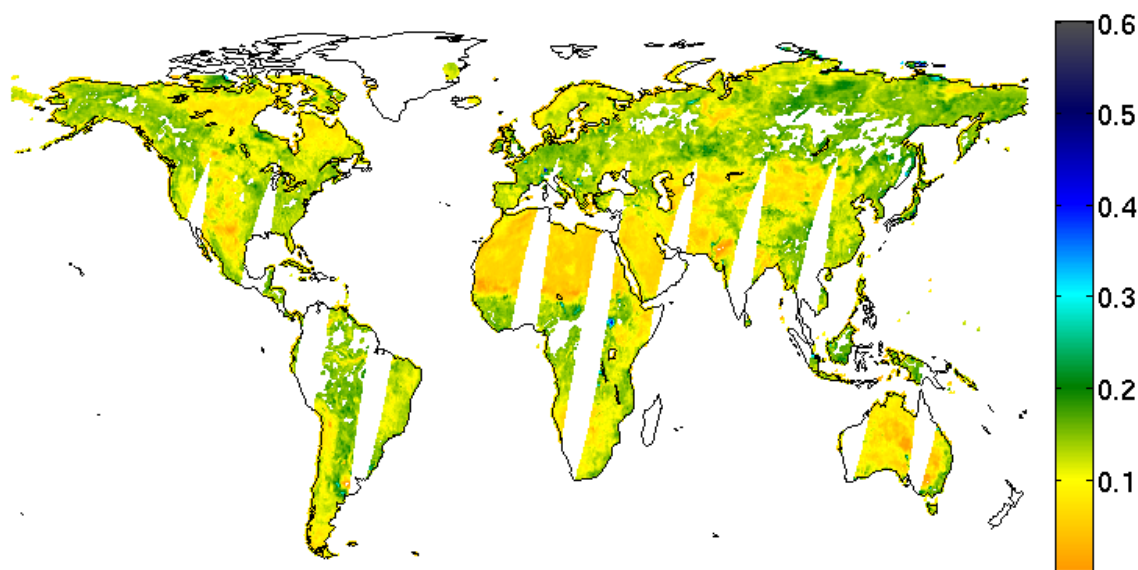


FIGURE 4.2 – Humidité des sols (m^3/m^3) retrouvée par NSIDC à partir des observations d'AMSR-E le 15 Septembre 2010 à 1:30h, heure solaire locale.

qui est important à cette fréquence. Comme les produits AMSR-E précédents, on estime que le signal est représentatif des 2 premiers centimètres du sol. Les données Level 2 ASCAT utilisées sont distribuées à deux résolutions (50 km et 25 km) et la résolution choisie pour cette étude est de 25 km.

Le produit humidité du modèle ECMWF utilisé dans cette partie est le produit qui sert de donnée auxiliaire à l'algorithme SMOS et qui a donc été interpolé spatialement et temporellement pour les besoins de cet algorithme. C'est un produit prévisionnel qui estime l'humidité (parmi beaucoup d'autres paramètres) entre 3 et 15 heures en avance et ceci toutes les 3 heures (3:00h, 6:00h, 9:00h, etc). Le produit ECMWF représente l'humidité des 7 premiers centimètres du sol.

Seules les données du matin sont utilisées : 1:30h pour les deux produits AMSR-E, 6:00h pour SMOS et ECMWF et 9:30h pour le produit ASCAT.

La journée du 15 Septembre 2010 illustre pour chaque jeu de données, le produit d'humidité des sols à l'échelle globale (figs. 4.1-4.5).

4.2 Distribution spatiale de l'erreur variable SMOS à l'échelle globale

4.2.1 Méthode de triple collocation

4.2.1.1 Principe

Puisque les mesures terrain ne sont pas réalisables à l'échelle globale, il est essentiel d'utiliser des méthodes statistiques pour comparer plusieurs jeux de données entre eux sur le globe. La triple collocation est une méthode statistique qui compare trois jeux de données et produit des cartes des erreurs relatives de chaque ensemble de données par rapport aux autres. Cette méthode a déjà été utilisée dans plusieurs domaines environnementaux : les vents à la surface de mer (Stoffelen, 1998, Guilfen et al., 2001, Caires & Sterl, 2003), la hauteur des vagues en mer (Caires & Sterl, 2003, Janssen et al., 2007), l'humidité des sols (Dorigo et al., 2010, Miralles

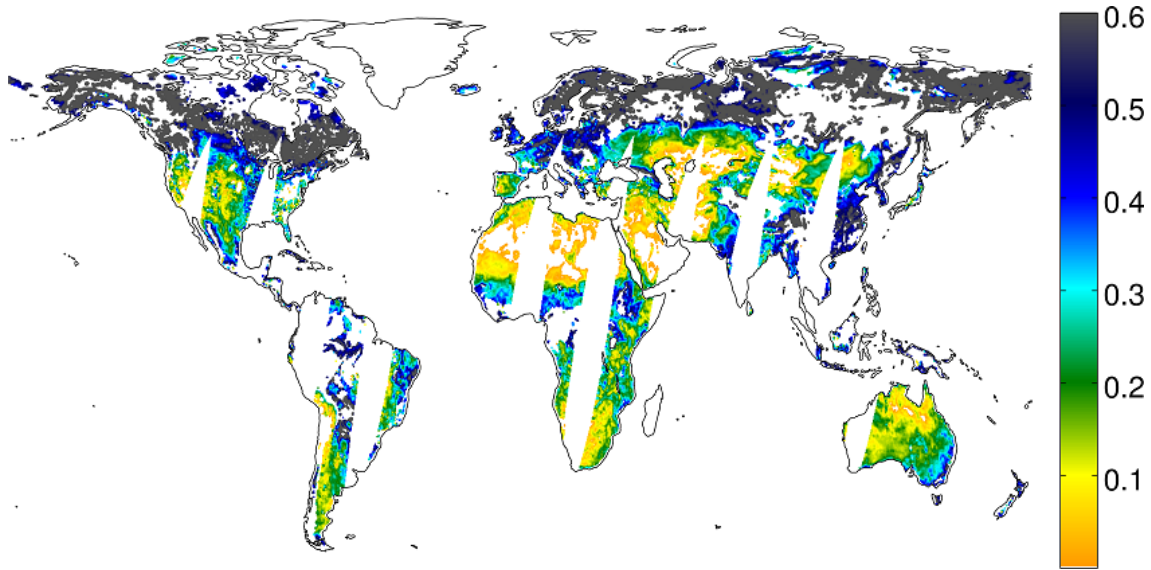


FIGURE 4.3 – Humidité des sols (m^3/m^3) retrouvée par VUA à partir des observations d'AMSR-E le 15 Septembre 2010 à 1:30h, heure solaire locale.

et al., 2010). Plus récemment, la triple collocation a été appliquée aux données d'humidité SMOS et à d'autres produits d'humidité de satellites actifs et de modèles sur une zone d'étude dans le Sud Ouest de la France (Parrens et al., 2011). A partir des 40 à 72 dates communes en 2010, il a été conclu que les humidités retrouvées à partir du capteur actif ASCAT donnaient de meilleurs résultats sur cette région avec une erreur relative de $0.031 \text{ m}^3/\text{m}^3$ alors que SMOS obtenait une erreur relative de $0.045 \text{ m}^3/\text{m}^3$. 72 dates considérées représentent cependant peu de données lorsque des statistiques sont calculées.

4.2.1.2 Théorie

Comme dans les études de Stoffelen (1998) et de Dorigo et al. (2010), l'approche proposée ici est une relation linéaire entre les trois jeux de données θ_i et la vérité supposée θ à l'aide d'un terme de biais r_i et d'un facteur d'échelle s_i . La méthode de triple collocation consiste à estimer les erreurs ε_i . Ces erreurs sont relatives à la vérité supposée θ et sont inter-comparables puisqu'elles sont relatives à la même quantité. Cependant, en aucun cas ces erreurs sont des erreurs absolues et ne peuvent être analysées à l'extérieur du triplet considéré. Un jeu de données est arbitrairement choisi comme référence de telle façon que $r_1 = 0$ et $s_1 = 1$. Ce choix de référence n'influe pas sur les résultats dans le sens où il ne sert que d'espace de projection pour les résultats. Choisir une autre référence aurait comme seule conséquence de changer la gamme des résultats finaux. Ainsi, on considère que θ a la même distribution que θ_1 mais il est important de remarquer que θ_1 n'est pas la vérité à cause du terme d'erreur ε_1 qui subsiste :

$$\begin{cases} \theta = r_1 + s_1\theta_1 + \varepsilon_1 \\ \theta = r_2 + s_2\theta_2 + \varepsilon_2 \\ \theta = r_3 + s_3\theta_3 + \varepsilon_3 \end{cases} \Rightarrow \begin{cases} \theta = \theta_1 + \varepsilon_1 \\ \theta = r_2 + s_2\theta_2 + \varepsilon_2 \\ \theta = r_3 + s_3\theta_3 + \varepsilon_3 \end{cases} \quad (4.1)$$

En prenant la moyenne sur l'année ($\langle . \rangle$) et en supposant que les erreurs ε_i ont une moyenne nulle, les expressions suivantes sont trouvées pour la moyenne de la vérité supposée :

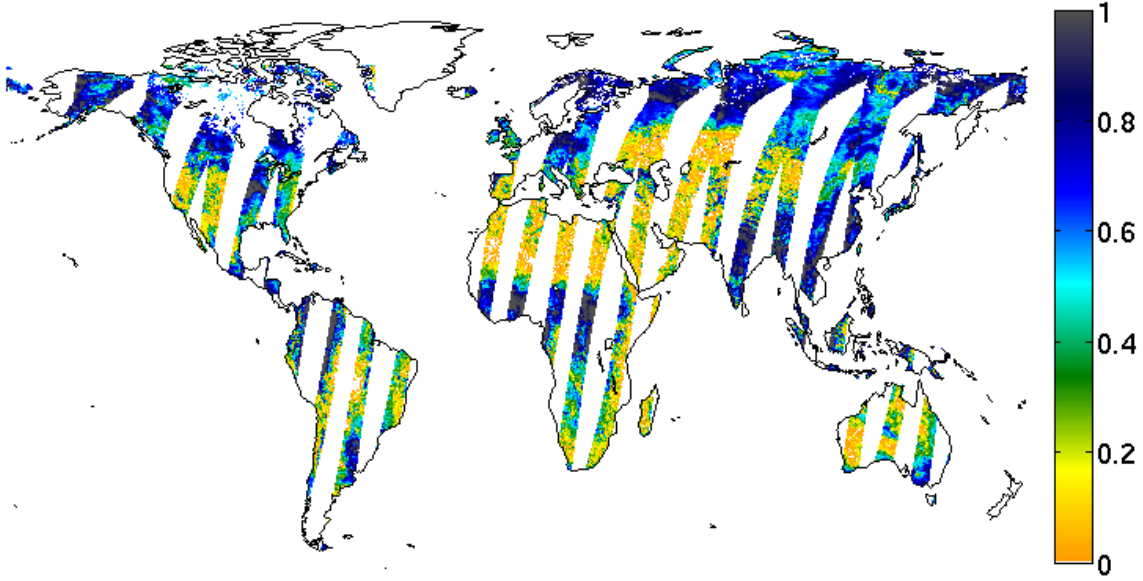


FIGURE 4.4 – Indice d'humidité des sols retrouvé par ASCAT le 15 Septembre 2010 à 9:30h, heure solaire locale.

$$\begin{cases} \langle \theta \rangle = & \langle \theta_1 \rangle \\ \langle \theta \rangle = & r_2 + s_2 \langle \theta_2 \rangle \\ \langle \theta \rangle = & r_3 + s_3 \langle \theta_3 \rangle \end{cases} \quad (4.2)$$

On définit θ'_i comme l'anomalie du jeu de données i par : $\theta'_i = \theta_i - \langle \theta_i \rangle$. En effectuant la soustraction de l'équation (4.1) avec l'équation (4.2), les termes de biais r_i disparaissent :

$$\begin{cases} \theta' = & \theta'_1 + \varepsilon_1 \\ \theta' = & s_2 \theta'_2 + \varepsilon_2 \\ \theta' = & s_3 \theta'_3 + \varepsilon_3 \end{cases} \quad (4.3)$$

Puis en multipliant les lignes entre elles, trois expressions de θ'^2 sont obtenues :

$$\begin{cases} \theta'^2 = & s_2 \theta'_1 \theta'_2 + s_2 \theta'_2 \varepsilon_1 + \theta'_1 \varepsilon_2 + \varepsilon_1 \varepsilon_2 \\ \theta'^2 = & s_3 \theta'_1 \theta'_3 + s_3 \theta'_3 \varepsilon_1 + \theta'_1 \varepsilon_3 + \varepsilon_1 \varepsilon_3 \\ \theta'^2 = & s_2 s_3 \theta'_2 \theta'_3 + s_2 \theta'_2 \varepsilon_3 + s_3 \theta'_3 \varepsilon_2 + \varepsilon_2 \varepsilon_3 \end{cases} \quad (4.4)$$

Les anomalies θ'_i sont supposées être indépendantes des erreurs des autres jeux de données ε_i et parce que les valeurs moyennes des variables θ'_i et ε_i sont nulles (θ'_i est l'anomalie à la moyenne et ε_i est un bruit additif de moyenne nulle), on obtient finalement $\langle \theta'_i \varepsilon_i \rangle = 0$.

En prenant les moyennes de chaque ligne de l'équation (4.4), on a :

$$\begin{cases} \langle \theta'^2 \rangle = & s_2 \langle \theta'_1 \theta'_2 \rangle \\ \langle \theta'^2 \rangle = & s_3 \langle \theta'_1 \theta'_3 \rangle \\ \langle \theta'^2 \rangle = & s_2 s_3 \langle \theta'_2 \theta'_3 \rangle \end{cases} \quad (4.5)$$

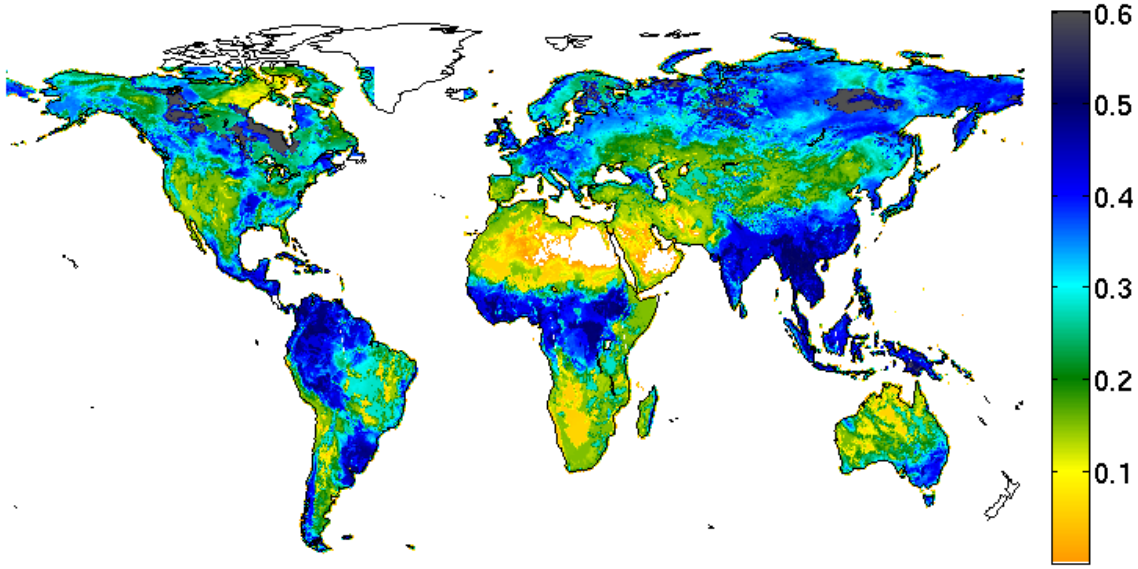


FIGURE 4.5 – Humidité des sols (m^3/m^3) retrouvée par le modèle ECMWF le 15 Septembre 2010 à 6:00h, heure solaire locale.

Et en combinant les lignes, on peut retrouver les facteurs d'échelle et la moyenne de la vérité supposée au carré :

$$\begin{cases} s_2 = \langle \theta'_1 \theta'_3 \rangle / \langle \theta'_2 \theta'_3 \rangle \\ s_3 = \langle \theta'_1 \theta'_2 \rangle / \langle \theta'_2 \theta'_3 \rangle \\ \langle \theta'^2 \rangle = \langle \theta'_1 \theta'_2 \rangle \langle \theta'_1 \theta'_3 \rangle / \langle \theta'_2 \theta'_3 \rangle \end{cases} \quad (4.6)$$

En manipulant les équations (4.2), on retrouve les expressions des termes de biais :

$$\begin{cases} r_2 = \langle \theta_1 \rangle - s_2 \langle \theta_2 \rangle \\ r_3 = \langle \theta_1 \rangle - s_3 \langle \theta_3 \rangle \end{cases} \quad (4.7)$$

En élevant au carré l'équation (4.3), en prenant sa moyenne et en remplaçant les termes de pente et de biais par leurs expressions définies par les équations (4.6) et (4.7), on obtient finalement les erreurs relatives moyennes suivantes :

$$\begin{cases} \langle \varepsilon_1^2 \rangle = \frac{\langle \theta'_1 \theta'_2 \rangle \langle \theta'_1 \theta'_3 \rangle}{\langle \theta'_2 \theta'_3 \rangle} - \langle \theta_1'^2 \rangle \\ \langle \varepsilon_2^2 \rangle = \frac{\langle \theta'_1 \theta'_2 \rangle \langle \theta'_1 \theta'_3 \rangle}{\langle \theta'_2 \theta'_3 \rangle} - \frac{\langle \theta'_1 \theta'_3 \rangle^2}{\langle \theta'_2 \theta'_3 \rangle^2} \langle \theta_2'^2 \rangle \\ \langle \varepsilon_3^2 \rangle = \frac{\langle \theta'_1 \theta'_2 \rangle \langle \theta'_1 \theta'_3 \rangle}{\langle \theta'_2 \theta'_3 \rangle} - \frac{\langle \theta'_1 \theta'_2 \rangle^2}{\langle \theta'_2 \theta'_3 \rangle^2} \langle \theta_3'^2 \rangle \end{cases} \quad (4.8)$$

Avec l'équation (4.8), il est ainsi possible de calculer les erreurs relatives en utilisant des éléments connus : les anomalies de chaque jeu de données par rapport à leur moyenne respective.

Puisque les erreurs ε_i ont des moyennes nulles, les quantités $\langle \varepsilon_i^2 \rangle$ peuvent être interprétées comme leurs variances. Le jeu de données avec la variance d'erreur la plus faible est alors considéré comme le plus précis des trois jeux de données testés. Le choix du jeu de données de

référence (la première ligne dans les équations) modifie les valeurs des erreurs relatives puisque l'espace de projection de ces erreurs change, mais il ne change pas la structure des erreurs : si le jeu de données n°2 obtient de mauvais résultats en Afrique comparé aux deux autres, il obtiendra de mauvais résultats même en étant choisi comme référence, mais avec des valeurs d'erreur différentes. Il est vraiment essentiel de comprendre que ces erreurs sont relatives au choix des trois jeux de données et au choix de la référence et elles ne peuvent pas être comparées à des erreurs trouvées à partir d'un autre triplet.

4.2.2 Conditions d'application et méthodologie

La triple collocation est une méthode basée sur les statistiques et ces statistiques n'ont de sens que si elles sont appliquées à un grand nombre de points. Dans le chapitre précédent lorsque plusieurs produits d'humidité ont été inter-comparés avec des mesures terrain, seulement quelques dizaines de points étaient disponibles, ce qui a engendré des intervalles de confiance très larges sur les coefficients de corrélation. S'il y avait eu deux ou trois fois plus de points, ces intervalles de confiance auraient été nettement réduits et cela aurait donné beaucoup plus de poids ou de valeur aux résultats statistiques obtenus. Scipal et al. (2010) ont déterminé qu'un minimum de 100 points est nécessaire pour appliquer la méthode de la triple collocation.

Dans cette étude, des produits d'humidité retrouvés à partir de données satellite et de modèle sont comparés sur l'année 2010 et en combinant les trois différentes orbites et fauchées, le minimum des 100 dates communes n'est pas atteint. Pour résoudre ce problème, les données des six plus proches voisins (qui sont distants de 15 km chacun d'un point central de la grille ISEA) sont utilisées comme si elles étaient les données du point central (fig. 4.6(a)). Sachant que (1) la grille est hexagonale, (2) la distance entre les points de la grille est de 15 km et (3) la résolution des instruments embarqués sur les satellites est d'environ 40 km (bandes L, C ou X), il est supposé que l'erreur introduite en considérant les six plus proches voisins est inférieure à l'erreur introduite par l'utilisation de moins de 100 dates communes par nœud de grille.

L'exemple proposé sur la figure 4.6 représente le nombre de dates en commun en Australie en 2010 entre les jeux de données SMOS, VUA et ECMWF. Le nombre de dates communes dépasse rarement la limite requise de 100 (fig. 4.6(b)). En prenant en compte les dates communes des six plus proches voisins, tous espacés de 15 km, constituant ainsi un disque de diamètre 30 km, seules les surfaces côtières ne dépassent pas les 100 dates communes (fig. 4.6(c)) et partout ailleurs sur ce continent, le nombre de dates disponibles oscille entre 100 et 800. Enfin, si un disque de diamètre de 60 km était considéré, ce qui correspondrait à prendre en compte les 18 proches voisins sur une grille hexagonale de 15 km, tout le continent serait admissible pour appliquer la méthode de triple collocation (fig. 4.6(d)). Compte tenu des caractéristiques des radiomètres considérés dont la résolution est de l'ordre de 30 km, il a été choisi de considérer seulement les six plus proches voisins. Cependant la triple collocation ne sera pas appliquée sur toutes les surfaces continentales puisque cette limite des 100 dates communes ne sera pas remplie partout.

Le biais étant considéré comme une plus grande déviation, il est important d'utiliser des produits d'humidité non biaisés entre eux quand on applique la triple collocation. Dans cette étude, la triple collocation est appliquée aux anomalies des produits par rapport à leur moyenne. Il n'existe donc aucun biais entre les produits car les anomalies ont une moyenne nulle.

Les erreurs sont supposées avoir une moyenne nulle et être mutuellement indépendantes de telle façon que leur covariance est nulle ($\langle \varepsilon_i \varepsilon_j \rangle = 0$). Autrement, il serait nécessaire d'ajouter un terme de cross-corrélation qui devrait être estimé. Afin d'éviter ce cas de dépendance entre les erreurs, les jeux de données testés sont précautionneusement choisis : les produits d'humidité retrouvés à partir des mêmes observations satellite (AMSR-E NSIDC et VUA) ne seront pas utilisés en même temps alors que les produits utilisant des observations et des algo-

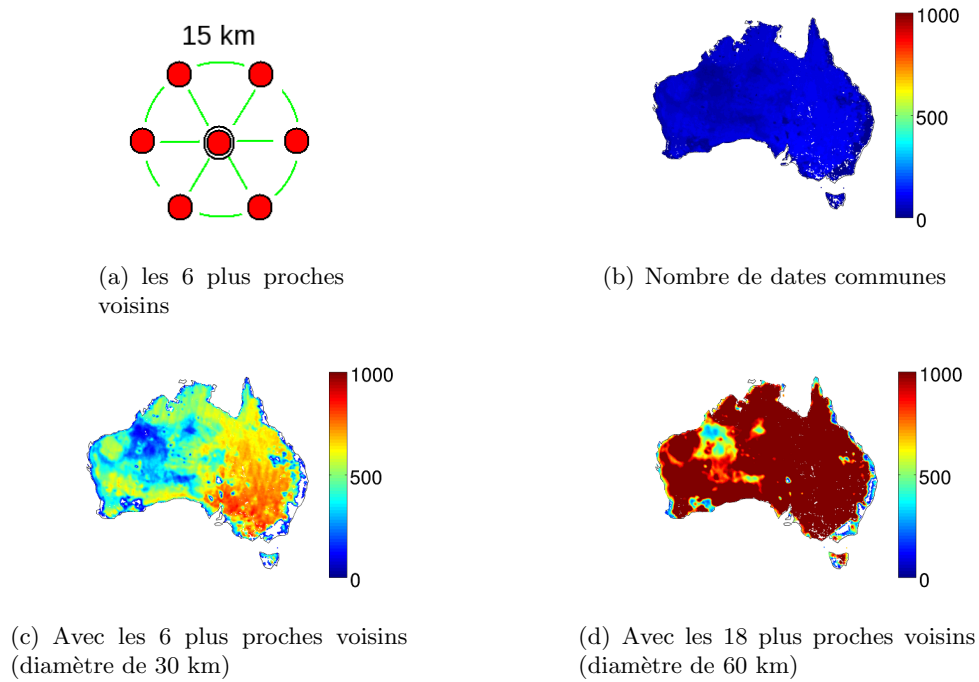


FIGURE 4.6 – La grille de SMOS est une grille hexagonale où tous les points sont espacés de 15 km (a). Exemple de l'Australie en 2010 avec les dates communes entre SMOS, VUA et ECMWF (b), puis en considérant les 6 plus proches voisins dans un rayon maximum de 30 km (c), et des 18 plus proches voisins dans un rayon maximum de 60 km.

rithmes différents pour retrouver l'humidité peuvent être comparés ensemble. Pour ces raisons, la méthode de triple collocation a été appliquée sur les triplets suivants : (SMOS - AMSR-E(VUA) - ECMWF), (SMOS - AMSR-E(NSDIC) - ASCAT) et (SMOS - AMSR-E(VUA) - ASCAT). Pour chacun de ces triplets, le produit d'humidité SMOS est considéré comme la référence.

La méthode de triple collocation peut être résumée en quatre étapes :

1. calculer les anomalies pour chaque point de la grille pour les trois jeux de données choisis,
2. calculer les facteurs d'échelle avec l'équation (4.6),
3. calculer les biais avec l'équation (4.7), et
4. calculer les variances moyennes des erreurs avec l'équation (4.8).

Les cartes d'erreur présentées dans la section suivante sont des cartes des variances moyennes des erreurs. Connaissant le nombre de dates communes qui a été utilisé pour calculer ces variances, il est possible de calculer l'intervalle de confiance associé à chaque variance d'erreur (voir annexe A.3). Ces intervalles de confiance permettent ainsi de connaître quel jeu de données obtient significativement l'erreur la plus faible : borne supérieure inférieure aux bornes inférieures des deux autres jeux de données.

De manière plus générale, la méthode de triple collocation se concentre sur les erreurs temporelles de chaque jeu de données (calcul des anomalies temporelles par rapport à la moyenne annuelle). Cette méthode permet donc d'identifier des erreurs temporelles et non spatiales. Cette différence pourrait favoriser les produits d'humidité qui ont été calculés à partir de méthode de type *detection change* comme ASCAT. L'algorithme du produit d'humidité ASCAT utilise les valeurs d'humidité des dernières décennies et compare les nouvelles observations avec ces

valeurs afin de détecter un changement et ainsi dériver une indice d'humidité.

4.2.3 Résultats

Dans cette section, les résultats suivants sont présentés : les cartes globales des erreurs relatives de chaque ensemble de données testé, les cartes de biais entre ces jeux de données et les cartes des facteurs d'échelle issus de la méthode de triple collocation. Les zones en blanc sur ces cartes sont les régions où le nombre minimum de dates communes de 100 pour appliquer la triple collocation n'est pas atteint et où aucune étude n'a été réalisée pour l'année 2010.

4.2.3.1 SMOS / AMSR-E(VUA) / ECMWF

La figure 4.7 montre les cartes d'erreurs des jeux de données SMOS, VUA et ECMWF. Au total, la triple collocation a été appliquée sur 302 474 points avec ce triplet.

SMOS a obtenu la plus faible erreur sur 17% des points (dont 10% significativement) alors que VUA est le meilleur sur 44% (dont 25% significativement) et ECMWF sur 39% (dont 19% significativement). Les erreurs SMOS les plus élevées se situent à l'Est des Etats-Unis, au Nord de l'Amérique du Nord, en Europe, en Inde et à l'Est de l'Asie. Ses meilleures performances sont localisées à l'Ouest des Etats-Unis, au Nord de l'Afrique, au Moyen-Orient, dans le centre de l'Asie et en Australie. ECMWF donne de très bons résultats en Europe, en Amérique du Sud, au centre des Etats-Unis et en Inde alors que les meilleurs résultats de VUA couvrent l'Ouest des Etats-Unis, certaines régions d'Afrique, en Asie et sur l'Ouest de l'Australie.

Il y a de nombreuses ressemblances dans les erreurs des deux produits d'humidité satellite. Les zones où les erreurs sont faibles et où elles sont élevées sont très similaires. SMOS et VUA n'ont pas de bons résultats dans les hautes latitudes, en Afrique centrale, à l'Est des Etats-Unis, sur les forêts d'Amérique du Sud ou encore en Asie du Sud-Est. Cela pourrait s'expliquer par le fait que ces régions sont différentes et leur modélisation en est plus complexe. Les régions en haute latitude sont souvent composées de lacs et de grandes étendues d'eau, ce qui influence fortement le signal émis. Même si une modélisation spécifique est dédiée à cette présence d'eau libre, cela représente tout de même une large proportion du signal et retrouver l'humidité sur les terres émergées devient moins précis. Les zones à forte densité de forêts sont également compliquées pour un instrument embarqué sur un satellite. En effet, en bande L, C ou X, le signal reçu par le radiomètre vient essentiellement des premiers centimètres du sol et lorsque la végétation est dense, le signal reçu ne vient probablement pas du tout du sol et cela est très compliqué à modéliser. En Asie, une forte présence d'interférences radio-fréquences (ou RFI) perturbent le signal, faisant augmenter artificiellement les températures de brillance mesurées par les radiomètres pouvant atteindre plusieurs milliers de kelvin. Ces mesures sont alors ininterprétables ou alors mal interprétées par le modèle qui retrouve des valeurs d'humidité très faibles.

Le modèle ECMWF se démarque par sa différence avec les produits d'humidité satellite : l'Inde, l'Afrique centrale ou encore l'Europe sont globalement bien modélisés avec une erreur plus faible. Les régions en haute latitude restent cependant difficilement retranscrites avec des erreurs élevées, même avec le modèle ECMWF.

Les facteurs d'échelle s_i sont calculés à partir des moyennes croisées des anomalies (Eq. 4.6). Ils représentent les facteurs par lesquels il faudrait multiplier les jeux de données pour qu'ils aient la même dynamique d'humidité, ou la même gamme de valeurs d'humidité, que la référence SMOS dans les trois cas. S'il n'existait aucun biais entre deux jeux de données, ce facteur d'échelle serait le ratio de leurs moyennes. Ces facteurs d'échelle sont exposés sur la figure 4.8.

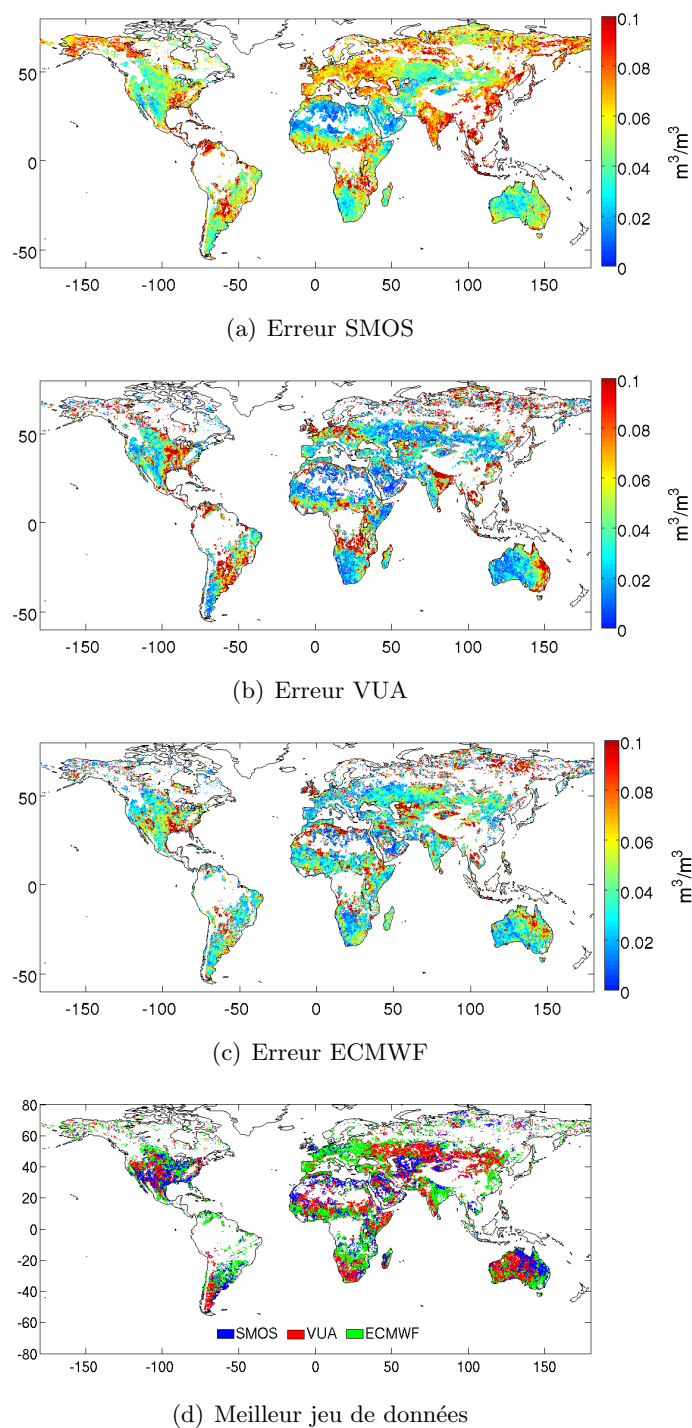
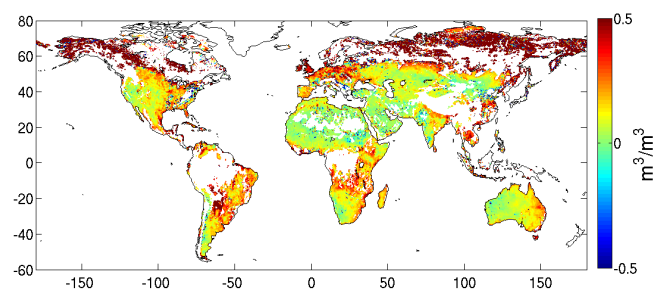
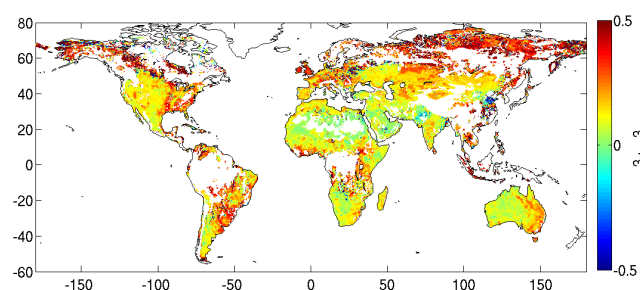


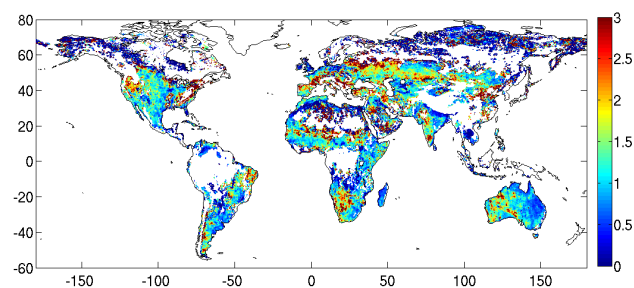
FIGURE 4.7 – Erreurs relatives de SMOS, VUA et ECMWF présentées à l'échelle globale et issues de la triple collocation appliquée en 2010 sur SMOS-VUA-ECMWF. Le meilleur jeu de données est celui dont l'erreur est significativement la plus basse.



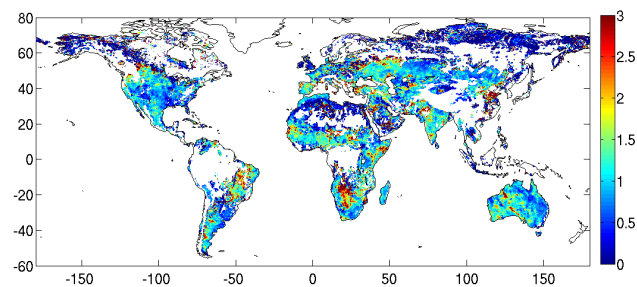
(a) Biais VUA



(b) Biais ECMWF



(c) Facteur d'échelle VUA



(d) Facteur d'échelle ECMWF

FIGURE 4.8 – Biais et facteurs d'échelle de VUA et d'ECMWF par rapport à SMOS, calculés à partir des équations (4.6) et (4.7).

Concernant le facteur d'échelle de VUA, on ne remarque pas de forme spatiale particulière, seulement des regroupements par tâches, excepté en Europe et en Asie centrale où une ligne traversant ces continents semblent se dessiner avec une valeur supérieure à 1 dénotant une plus petite dynamique de la part de VUA par rapport à SMOS. VUA présente cependant une plus grande dynamique que SMOS dans les régions à haute latitude et dans l'Est de l'Australie. Les facteurs d'échelle d'ECMWF par rapport à SMOS ne suivent pas non plus de distribution spatiale particulière. On remarque cependant moins de valeurs élevées que pour le produit VUA.

Après que les facteurs d'échelle aient été déterminés, les biais sont calculés (Eq. 4.7) en soustrayant la moyenne des humidités de SMOS avec celle des humidités d'un second produit corrigé de son facteur d'échelle. Les biais de VUA et d'ECMWF sont montrés sur la figure 4.8. Le biais VUA est quasiment totalement positif avec des valeurs autour de $0.5 \text{ m}^3/\text{m}^3$ aux hautes latitudes. Ce biais élevé est très certainement dû au facteur d'échelle qui est très faible dans ces régions. Les biais sont plus hauts sur l'Europe, l'Amérique du Sud et l'Afrique centrale alors qu'ils sont quasi nuls en Afrique du Nord, au moyen Orient et dans l'Ouest de l'Australie, régions plutôt désertiques avec très peu de végétation.

Le biais ECMWF est plus homogène dans sa distribution spatiale même s'il reste plus élevé dans les hautes latitudes et moins élevé dans les régions désertiques également. Excepté la petite région à l'extrême Est de la Chine, le biais du produit ECMWF est positif après correction du facteur d'échelle comme le biais de VUA.

4.2.3.2 SMOS / AMSR-E(NSIDC) / ASCAT

Les cartes d'erreur des produits SMOS, NSIDC et ASCAT sont présentées sur la figure 4.9. La triple collocation a été appliquée sur 247 798 points, ce qui représente une diminution de 20% de la surface étudiée qui est dû au fait que dans ce triplet, trois produits d'humidité dérivés d'observations satellites sont comparés alors que précédemment deux produits satellite et un produit modèle étaient comparés. Le modèle ECMWF retrouve l'humidité partout et n'a pas de problème de largeur de fauchée alors qu'ASCAT étant un satellite, le nombre de points disponibles pour la comparaison est moins élevé.

SMOS fournit des erreurs plus basses sur 21% des points (dont 14% significativement) contre 35% (dont 7% significativement) pour NSIDC et 44% (dont 19% significativement) pour ASCAT. Avec ce triplet, on remarque que les produits ASCAT et NSIDC sont les meilleurs mais finalement pas de manière significative comparés au produit SMOS. Les meilleures performances de SMOS se situent en Afrique du Nord, au Moyen-Orient, à l'Ouest des Etats-Unis et à l'Ouest de l'Australie, ce qui correspond aux zones de prédilection de SMOS avec le triplet précédent. En revanche, SMOS obtient une forte erreur en Europe, à l'Est de l'Amérique du Sud, en Afrique centrale et dans les régions aux hautes latitudes. Le produit NSIDC obtient des erreurs très élevées quasiment partout sauf en Afrique du Nord sur le désert du Sahara. Quant aux erreurs ASCAT, elles sont plus distribuées et ne sont pas plus élevées aux hautes latitudes. Elles restent cependant fortes aux Etats-Unis, en Europe et en Australie. On peut noter également que l'erreur est plus forte dans la zone de transition en Afrique où les forêts sont peu denses entre le désert aride du Sahara et les régions chaudes et humides près de l'équateur où les forêts sont très denses.

La carte de facteur d'échelle du produit NSIDC (fig. 4.10) montre des valeurs extrêmement faibles et inférieures à 1. Cela veut dire que le produit NSIDC utilise une gamme de valeurs d'humidité beaucoup plus large que SMOS et qu'un facteur d'échelle très petit est nécessaire pour faire correspondre ces largeurs de gamme. Si SMOS décrit une humidité entre 0.1 et $0.2 \text{ m}^3/\text{m}^3$ et que dans un même temps NSIDC donne des valeurs entre 0.2 et $0.4 \text{ m}^3/\text{m}^3$ un facteur d'échelle de 0.5 est nécessaire pour que les deux largeurs de gammes se correspondent. Ces facteurs d'échelle très faibles dénotent donc un *sur*-dynamisme du produit NSIDC par rapport

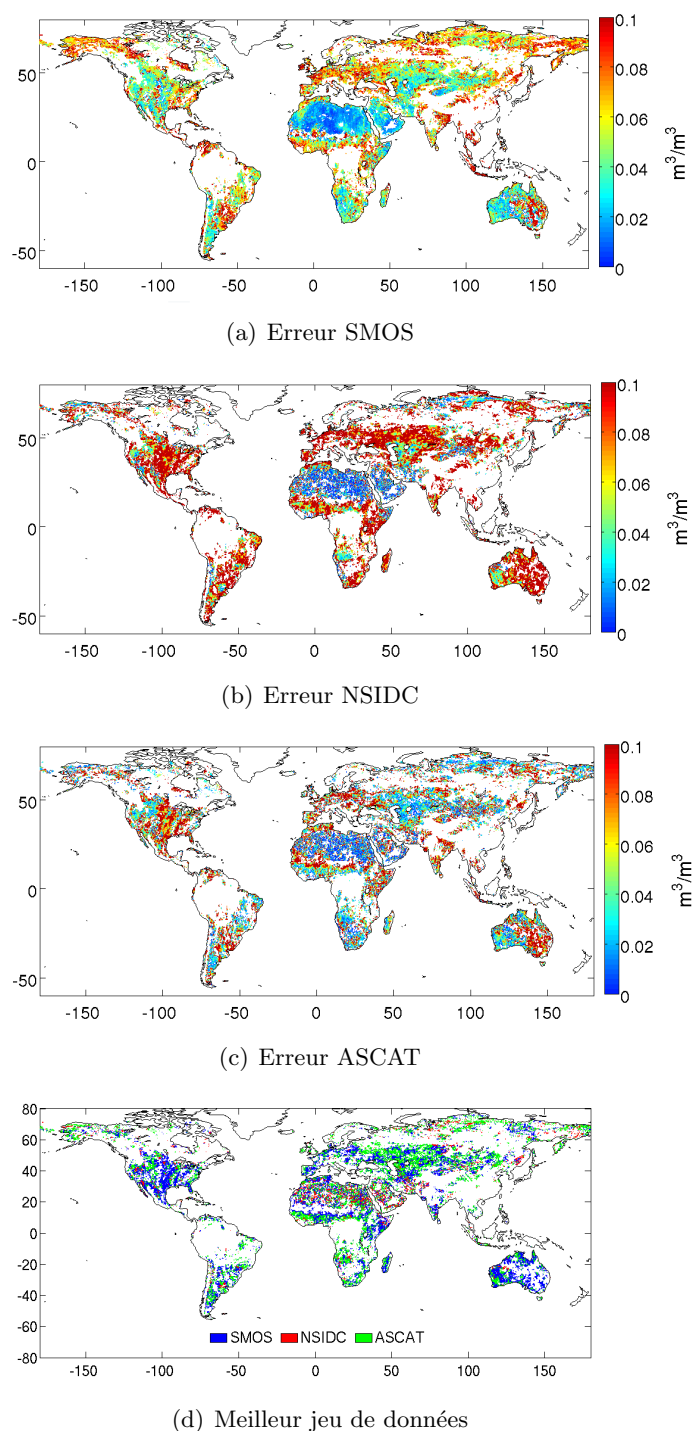
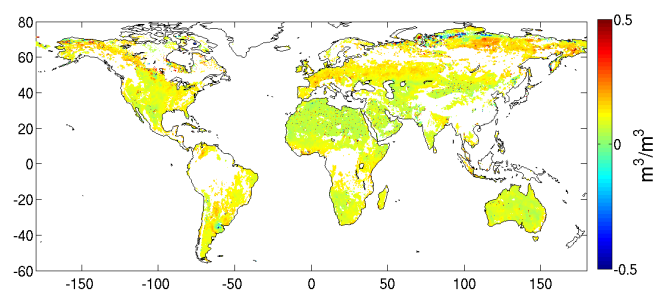
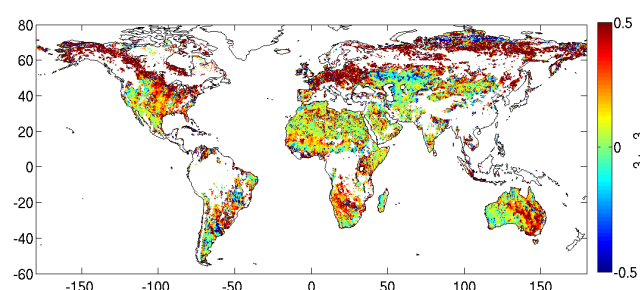


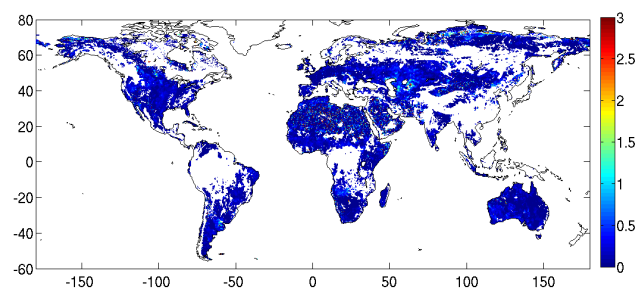
FIGURE 4.9 – Erreurs relatives de SMOS, NSIDC et ASCAT présentées à l'échelle globale et issues de la triple collocation appliquée en 2010 sur SMOS-NSIDC-ASCAT. Le meilleur jeu de données est celui dont l'erreur est significativement la plus basse.



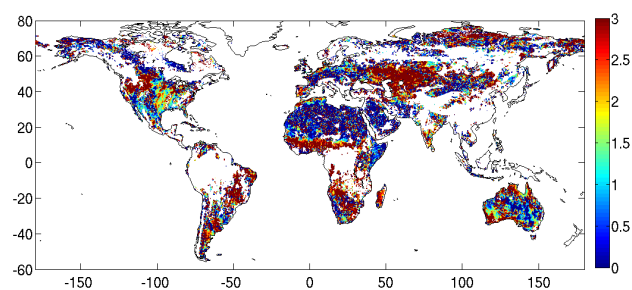
(a) Biais NSIDC



(b) Biais ASCAT



(c) Facteur d'échelle NSIDC



(d) Facteur d'échelle ASCAT

FIGURE 4.10 – Biais et facteurs d'échelle de NSIDC et d'ASCAT par rapport à SMOS, calculés à partir des équations (4.6) et (4.7).

au produit SMOS. ASCAT en revanche présente une distribution spatiale du facteur d'échelle intéressante avec soit une valeur très élevée, soit une valeur très faible mais pas de valeur intermédiaire. On peut constater ainsi qu'en Europe, en Afrique du Nord, au moyen Orient, les facteurs d'échelle sont très bas et qu'en Asie centrale, en Afrique centrale, en Amérique du Sud, ils sont très élevés.

Le biais entre SMOS et NSIDC est très stable partout sur le globe (entre 0 et $0.1 \text{ m}^3/\text{m}^3$). Il est quasi nul sur les régions à faible végétation : Afrique du Nord, moyen Orient, Australie. Le biais ASCAT évolue beaucoup plus et atteint $0.5 \text{ m}^3/\text{m}^3$ sur les régions aux hautes latitudes mais reste faible partout ailleurs malgré quelques hausses en Australie notamment.

4.2.3.3 SMOS / AMSR-E(VUA) / ASCAT

La figure 4.11 montre les cartes des erreurs SMOS, VUA et ASCAT où la triple collocation a été appliquée sur 210 368 points au total. Moins de points sont disponibles qu'avec le triplet précédent alors que seul le produit NSIDC a été remplacé par le produit VUA. Même si ces deux produits utilisent les mêmes observations d'AMSR-E pour retrouver l'humidité, leurs algorithmes sont très différents et VUA retrouve moins souvent une valeur d'humidité que NSIDC.

SMOS possède une erreur plus faible sur 21% des points (dont 12% significativement) alors que VUA a une meilleure performance sur 34% (15% significativement) et ASCAT sur 45% (dont 19% significativement). Avec ce triplet, on peut noter que les trois produits sont équivalents car ils obtiennent à peu près le même nombre de points où ils sont significativement meilleurs. Comme cela a été le cas précédemment, SMOS obtient de meilleurs résultats en Afrique du Nord, au moyen Orient, à l'Ouest des Etats-Unis, en Asie centrale et en Australie; et a de moins bonnes performances en Europe, sur les régions aux hautes latitudes, en Asie, en Afrique centrale. Le produit VUA présente de faibles erreurs en Afrique du Nord, en Asie centrale et à l'Ouest des Etats-Unis et a de moins bonnes performances à l'Est des Etats-Unis, en Europe au Sud de l'Amérique du Sud et en Australie. La carte des erreurs ASCAT est relativement homogène sauf aux Etats-Unis et au Sud de l'Amérique du Sud où les erreurs sont plus élevées.

La carte du facteur d'échelle du produit VUA par rapport à SMOS (fig. 4.12) ne suit pas particulièrement de forme spatiale intéressante. On peut noter tout de même l'homogénéité de ce facteur sur l'Australie qui oscille autour de 1 alors que partout ailleurs, il évolue très rapidement spatialement. Le facteur d'échelle ASCAT ressemble à celui du triplet précédent avec des valeurs qui sont soit très élevées, soit très faibles. Les valeurs hautes se trouvent sur toute l'Europe et l'Asie centrale, sur les Etats-Unis, l'Amérique du Sud et l'Afrique centrale alors que les valeurs basses sont sur l'Afrique du Nord et le moyen Orient où il n'y a quasiment aucune végétation.

Concernant les cartes de biais (fig. 4.12), on retrouve ce biais positif de $0.5 \text{ m}^3/\text{m}^3$ sur les hautes latitudes aussi bien pour VUA que pour ASCAT. Le biais VUA reste néanmoins assez stable et bas après une zone de transition avec ces régions à haute latitude. Le biais ASCAT n'est pas aussi homogène et peut beaucoup varier spatialement.

4.2.3.4 Distribution spatiale de l'erreur variable de SMOS

Le calcul de l'erreur globale SMOS à travers la triple collocation a été effectué en comparant le produit SMOS à d'autres produits satellite et modèle. Les trois essais de triple collocation sur trois triplets différents concordent quant à la distribution de l'erreur SMOS montrant de meilleures performances en Amérique du Nord, en Afrique du Nord, au Moyen-Orient, en Asie centrale et en Australie. En général, SMOS donne de meilleurs résultats en Amérique du Nord, dans l'ouest de l'Asie et en Australie. Ces régions en particulier n'ont pas de source majeure

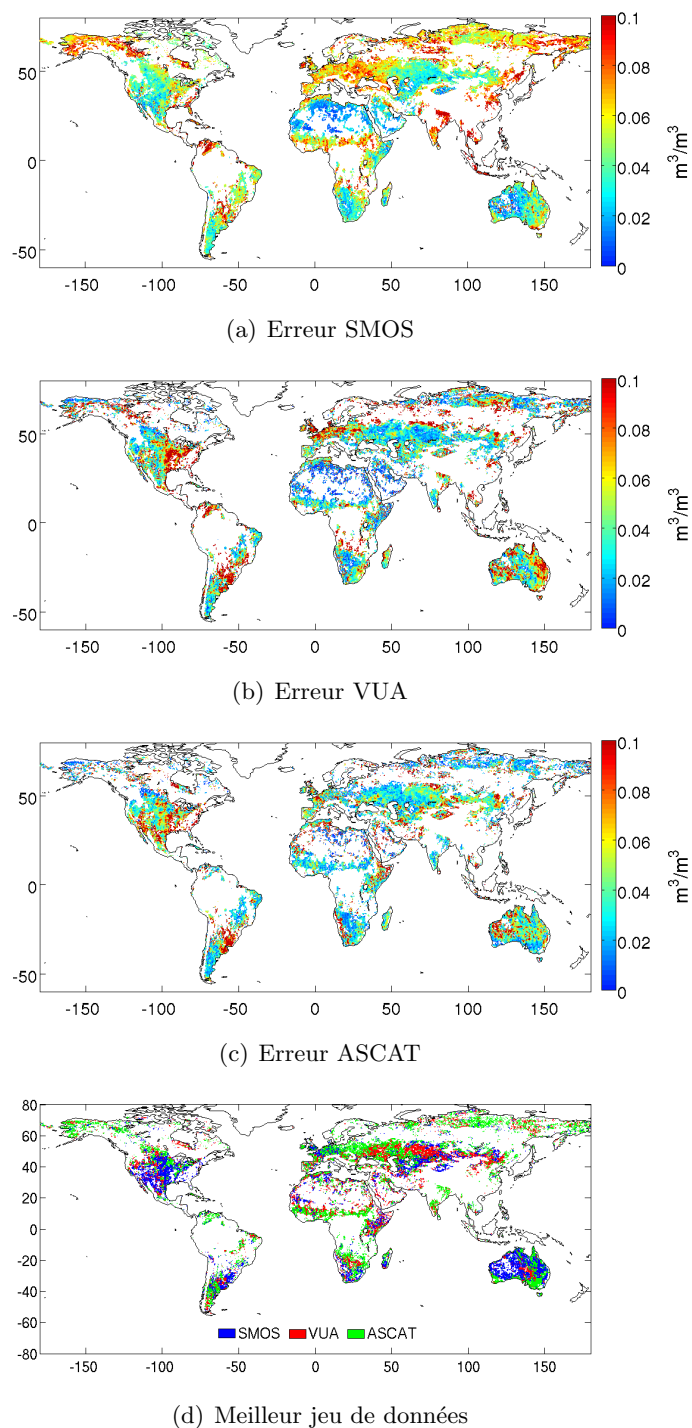
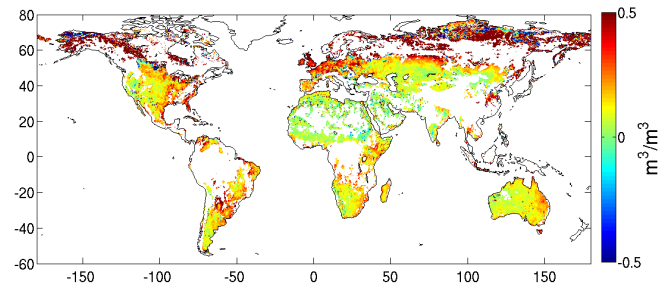
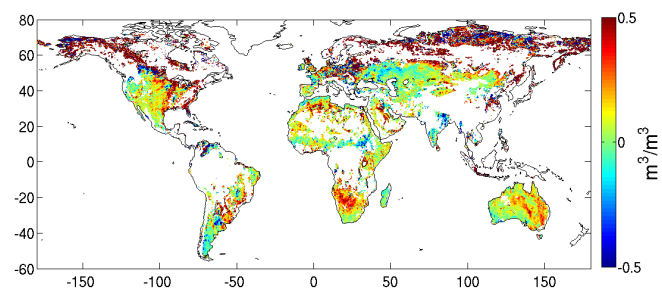


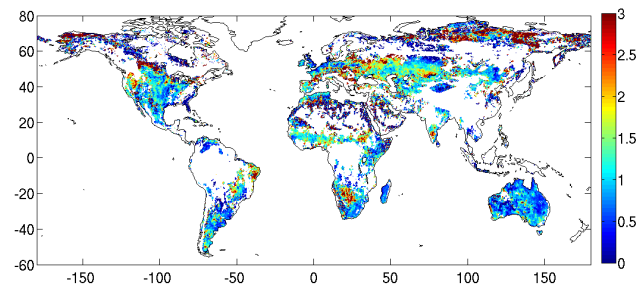
FIGURE 4.11 – Erreurs relatives de SMOS, VUA et ASCAT présentées à l'échelle globale et issues de la triple collocation appliquée en 2010 sur SMOS-VUA-ASCAT. Le meilleur jeu de données est celui dont l'erreur est significativement la plus basse.



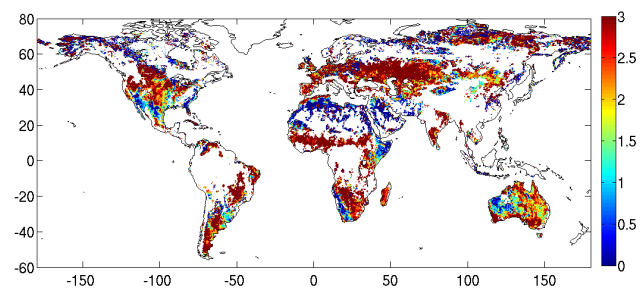
(a) Biais VUA



(b) Biais ASCAT



(c) Facteur d'échelle VUA



(d) Facteur d'échelle ASCAT

FIGURE 4.12 – Les biais et facteurs d'échelle de VUA et d'ASCAT par rapport à SMOS. Ils ont été calculés à partir des équations (4.6) et (4.7).

identifiée qui brouillerait le signal reçu par l'instrument SMOS, ce qui pourrait être une explication de cette localisation.

Bien que les erreurs des trois triplets précédents aient été calculées avec le même jeu de données de référence, elles ne peuvent pas être comparées car les dates communes ne sont pas les mêmes pour trois triplets. En revanche, leurs distributions spatiales sont très similaires sur les trois cartes d'erreur SMOS ainsi générées (fig. 4.7, 4.9 et 4.11). Les régions où SMOS donne de bons résultats sont identiques pour les trois triplets testés. On pouvait s'interroger sur la possible corrélation entre les erreurs SMOS et les erreurs ECMWF (les données ECMWF sont utilisées dans l'algorithme SMOS donc si ECMWF présente une erreur, elle va se propager) mais les cartes d'erreur montrent bien que la distribution générale de l'erreur SMOS est la même pour les trois triplets. Ainsi, ces cartes d'erreur peuvent être utilisées pour comprendre quelles sont les sources de ces erreurs et répondre à ce type de question : *pourquoi SMOS a une erreur plus faible en Australie qu'en Europe ?* Il n'existe pas de réponse directe mais très certainement qu'un ensemble de circonstances a amené l'algorithme SMOS à retrouver une humidité plus ou moins précise : comme la texture du sol, l'occupation du sol ou la présence d'interférences.

4.3 Analyse régionale et globale de l'erreur variable de SMOS

La distribution spatiale à l'échelle globale des erreurs variables de SMOS a été déterminée par triple collocation. À l'aide de deux méthodes statistiques basées sur l'analyse de la variance et sur une classification, ces erreurs vont être reliées à des paramètres physiques ou algorithmiques. L'analyse de la variance permettra de déterminer quel paramètre est responsable de quelle proportion d'erreur et la classification permettra de mettre en avant des ensembles de valeurs de ces paramètres qui conduisent à de très fortes ou de très faibles erreurs. Ces paramètres pouvant varier fortement régionalement, une étude globale et continentale est menée.

4.3.1 Présentation des paramètres physiques et algorithmiques possiblement liés à l'erreur variable de SMOS

Afin de retrouver les principales sources des erreurs SMOS, les paramètres géophysiques suivants ont été étudiés :

- le pourcentage de sable dans le sol (% sable),
- le pourcentage d'argile (% argile),
- la probabilité d'occurrence d'interférences en 2010 (RFI),
- la fraction de forêt vue par l'instrument SMOS (FFO),
- la fraction de zone humide (FWL),
- la fraction d'eau pure (FWP),
- la fraction d'eau salée (FWS),
- la fraction de sol nu ou aride (FEB),
- la fraction de glace (FEI) et
- la fraction de zone urbaine (FEU).

Ce sont les fractions que l'instrument SMOS est supposé voir et elles ont été déterminées par les classes ECOCLIMAP et peuvent ne pas correspondre à la réalité mais c'est pour l'instant la meilleure estimation disponible à l'échelle globale de ce que voit l'instrument. La fraction de végétation basse n'a pas été ajoutée à l'étude car elle correspond à une combinaison d'autres fractions (leur somme est égale à 1) et si elle avait été ajoutée, elle serait devenue un paramètre de contrainte. Le but de cette étude est avant tout de trouver les ensembles de paramètres qui détériorent les performances de SMOS et les zones à faible végétation n'en font pas partie. De plus, les paramètres de gel et de neige n'ont pas été pris en compte car ce sont des paramètres

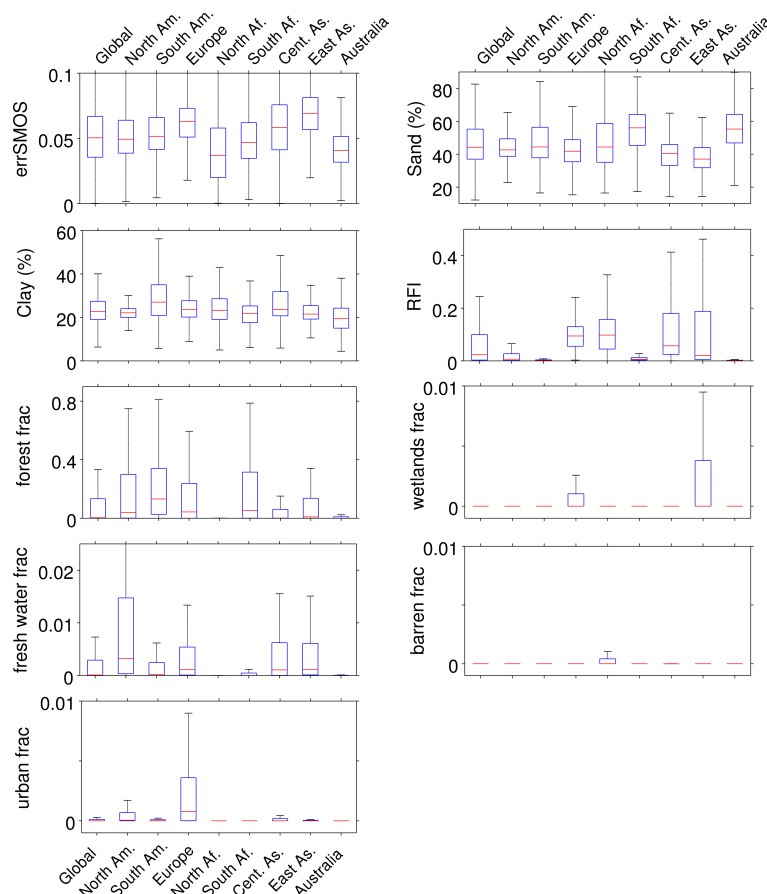


FIGURE 4.13 – Médiane (trait rouge), 1er et 3ème quartile (boîte bleue) de chaque paramètre étudié à l'échelle globale et continentale en 2010 (à l'exception des fraction d'eau salée et de glace car elles étaient nulles ou très proches de 0 pour tous les continents). L'erreur SMOS est issue de la triple collocation entre SMOS, VUA et ECMWF.

qui évoluent dans le temps et les erreurs SMOS sont ici fixes pour l'année 2010 donc elles doivent être comparées à des éléments fixes tels que l'occupation du sol. Une probabilité d'occurrence comme les RFI aurait pu être envisagée pour les événements de neige et de gel.

La carte des erreurs SMOS retrouvées à partir du triplet SMOS - VUA - ECMWF est choisie pour cette étude car c'est le triplet qui comporte le plus de points et qui couvre donc plus de surface. Il est ainsi plus représentatif.

Les valeurs moyennes et les écart-types de l'erreur variable de SMOS et des paramètres étudiés ont été calculés pour l'année 2010 et sont montrés en figure 4.13. Comme précédemment remarqué sur la carte des erreurs SMOS (fig. 4.7), l'Afrique du Nord et l'Australie sont les deux continents où l'erreur SMOS est la plus faible : $0.040 \text{ m}^3/\text{m}^3$ et $0.043 \text{ m}^3/\text{m}^3$. L'erreur est cependant plus homogène en Australie avec un écart-type plus faible ($0.015 \text{ m}^3/\text{m}^3$).

Le pourcentage d'argile est relativement stable sur tous les continents (autour de 24%). L'Amérique du Sud est le continent où ce paramètre est le plus hétérogène avec un grand écart-type de 10% alors que l'Asie de l'Est est la plus homogène avec un écart-type de 5%. Le pourcentage de sable est bien plus hétérogène que l'argile. Les sols d'Afrique du Sud et d'Australie sont ceux où le taux de sable est le plus élevé (55.9% et 55.6%) alors que l'Asie de l'Est est encore la région la plus hétérogène avec un écart-type de 9%. Les interférences radio-fréquences n'affectent pas tous les continents de manière égale. Alors que l'Australie,

Tableau 4.1 – Coefficients de corrélation de la régression linéaire multiple à l'échelle globale puis continentale.

	Global	Am. Nord	Am. Sud	Europe	Af. Nord	Af. Sud	Asie Cent.	Asie Est	Australie
R_{tot}	0.455	0.425	0.470	0.540	0.387	0.606	0.523	0.451	0.610

l'Amérique et l'Afrique du Sud ne sont quasiment pas concernés, l'Europe, l'Afrique du Nord et l'Asie sont hautement contaminés par ces interférences. De plus, ces statistiques ne concernent que les points où une valeur d'humidité a été retrouvée et il existe un très grand nombre de points où les interférences sont tellement fortes qu'aucune valeur d'humidité ne peut être retrouvée. Si on avait considéré ces points, les statistiques concernant les interférences auraient été bien plus élevées. La fraction de forêt estimée par ECOCLIMAP (et que l'on suppose vue par l'instrument SMOS) est très hétérogène, plus spécialement en Afrique du Nord et en Australie. Les autres fractions ne représentent pas une grande proportion de ce qui est vu par le radiomètre. Cependant, il peut être noté qu'il y a plus de zones humides en Europe et en Asie de l'Est, plus d'eau libre en Amérique du Nord, plus de sol nu ou aride en Afrique du Nord et plus de zones urbaines en Europe et en Amérique du Nord.

4.3.2 Analyse de la variance (ANOVA)

4.3.2.1 Principe et théorie

L'analyse de la variance (ANOVA) est une méthode statistique s'appuyant sur les évolutions de la variance de chaque paramètre qui permet de déterminer quel paramètre est responsable de quelle proportion d'erreur. Plus précisément, cette méthode étudie la relation entre une variable Y et plusieurs variables explicatives X_1, X_2, \dots, X_n . Cette méthode n'est représentative que de la forme linéaire suivante (où δ est le terme d'erreur qui reste non expliqué) :

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n + \delta \quad (4.9)$$

$$\begin{aligned} \langle \varepsilon_{smos}^2 \rangle = & \alpha_o + \alpha_{clay} X_{clay} + \alpha_{sand} X_{sand} + \alpha_{RFI} X_{RFI} + \\ & \alpha_{FFO} X_{FFO} + \alpha_{FWL} X_{FWL} + \alpha_{FWP} X_{FWP} + \alpha_{FWS} X_{FWS} + \\ & \alpha_{FEB} X_{FEB} + \alpha_{FEI} X_{FEI} + \alpha_{FEU} X_{FEU} + \delta \end{aligned} \quad (4.10)$$

Si les variables explicatives n'ont pas la même unité que la variable à expliquer Y , ce qui est le cas dans cette étude, il est absolument nécessaire de normaliser chaque variable avant de l'utiliser dans cette relation linéaire. Autrement, cette équation n'a pas de sens. Tous les paramètres α sont ensuite calculés avec la méthode classique des moindres carrés.

L'ANOVA est une méthode statistique utilisée pour étudier la modification de la moyenne d'une variable selon l'influence d'une ou de plusieurs variables explicatives. Les coefficients α_i sont alors les fractions de variance de Y expliquées par les variables X_i . Dans cette étude, on cherche à expliquer la distribution de l'erreur variable de SMOS à l'aide de plusieurs paramètres et on dira qu'un paramètre a une influence positive lorsque son coefficient α est négatif, i.e. une augmentation de la valeur de ce paramètre entraîne une diminution de l'erreur.

4.3.2.2 Résultats

Puisque l'on dispose de dix variables explicatives X et afin de ne garder que les variables qui expliquent réellement Y , les variables expliquant moins d'1% de la variance de l'erreur

Tableau 4.2 – Pourcentages de la variance de l'erreur SMOS expliqués par chaque variable explicative : pourcentages de sable et d'argile, probabilité moyenne de RFI sur 2010 (RFI), fraction de forêt (FFO), fraction de zone humide (FWL), fraction d'eau libre (FWP), fraction d'eau salée (FWS), fraction de sol aride (FEB), fraction de surface gelée (FEI) et fraction de zone urbaine (FEU).

	% sable	% argile	RFI	FFO	FWL	FWP	FWS	FEB	FEI	FEU
Global	22.2	0.1	1.7	56.8	7.2	3.3	2.4	2.0	0.1	4.2
Am. Nord	7.8	2.6	0.6	66.9	3.5	4.0	4.5	2.5	0.6	7.0
Am. Sud	39.3	1.4	6.2	42.8	0.1	4.8	2.5	1.2	0.1	1.6
Europe	3.5	42.7	24.4	7.9	9.3	2.4	4.2	0.2	0.2	5.2
Af. Nord	2.3	14.7	18.1	50.2	3.4	4.7	1.2	0.1	-	5.3
Af. Sud	8.1	37.9	2.8	48.2	0.6	1.4	0.3	0.5	-	0.2
Asia cent.	0.8	5.3	47.0	34.7	3.1	3.4	4.7	0.3	0.2	0.5
Asia Est	9.2	29.7	15.6	9.7	15.8	2.9	3.4	13.2	0.0	0.5
Australie	52.5	1.9	11.8	29.5	0.2	2.4	1.3	0.3	-	0.1

SMOS d'après l'ANOVA sont retirées du modèle de régression. Les coefficients de corrélations du modèle de régression linéaire sont indiqués dans le tableau 4.1. Les valeurs de corrélation peuvent paraître faibles mais ils sont suffisants compte tenu du nombre de variables explicatives pour poursuivre l'analyse.

L'influence de chaque paramètre explicatif sur la variance de l'erreur SMOS est présentée sous forme de diagramme sur la figure 4.14 et les proportions exactes sont indiquées dans le tableau 4.2.

À l'échelle globale, la régression linéaire multiple suivie de l'ANOVA montre que plus de la moitié (57%) de la variation de l'erreur SMOS est due à la variation de la fraction de forêt, avec une influence négative, c'est à dire que plus le radiomètre voit de la forêt, plus l'erreur SMOS est grande. La deuxième variable explicative la plus importante est le pourcentage de sable dans le sol, représentant 22% de la variation de l'erreur SMOS avec une influence négative également. Le troisième paramètre le plus influent est la fraction de zones humides avec 7% en négatif. En plus de représenter 86% de l'erreur SMOS, il est important de noter que si l'un de ces trois paramètres est mal estimé cela aura un impact non négligeable sur l'erreur SMOS finale.

Les RFI n'ont pas une grande influence à l'échelle globale sur l'erreur SMOS (moins de 2%). Cependant, son influence peut être très grande à l'échelle régionale.

La seule influence positive à l'échelle globale est la fraction de sol nu ou aride (FEB), i.e., plus il y a de sol nu, moins l'erreur est grande. Ces résultats étaient globalement attendus puisque les fractions d'eau, de zones humides, de zones urbaines ou de forêt perturbent le signal et rendent plus compliquée la modélisation de l'humidité du sol sur les zones considérées.

À l'échelle continentale, de nombreuses différences peuvent être identifiées. La fraction de forêt (FFO) qui est le facteur le plus influent à l'échelle globale, est toujours la plus influente sur la majorité des continents mais son influence ne représente que 10% en Europe et Asie de l'Est alors qu'elle représente au moins 30% sur les autres continents. Il est possible que le modèle de forêt implémenté dans l'algorithme SMOS soit plus adapté aux forêts européennes.

Sur l'Amérique du Nord, la fraction de forêt représente 67% de la variation de l'erreur SMOS. En effet, les erreurs SMOS ont été identifiées comme étant plus importantes sur les hautes latitudes et ce sont sur ces régions que les forêts se densifient (fig. 4.7). Les deuxième et troisième paramètres les plus influents sont la fraction de zone urbaine et le pourcentage de sable dans le sol avec 8% et 7% respectivement. Sur ce continent, l'influence positive du

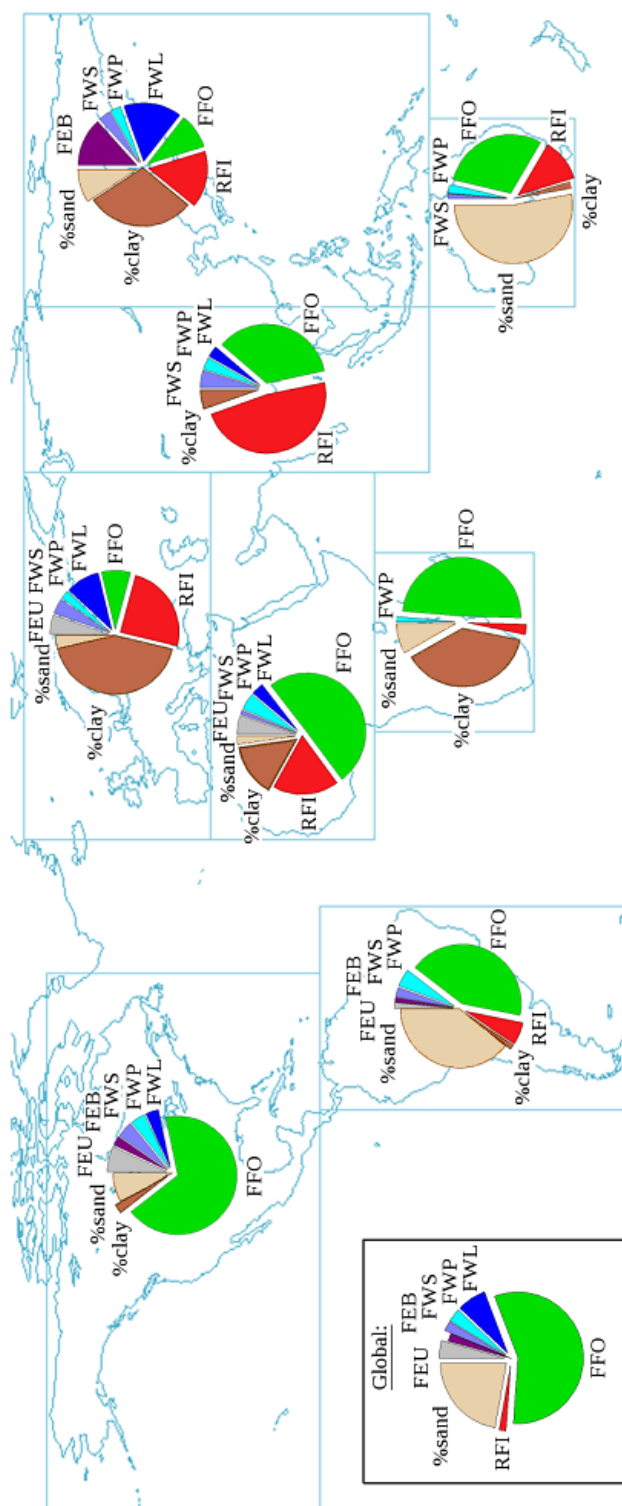


FIGURE 4.14 – Influence de chaque paramètre sur l'erreur SMOS à l'échelle globale et continentale : pourcentage de sable et d'argile (% sand, % clay), probabilité moyenne d'occurrence de RFI en 2010 (RFI), fraction de forêt (FFO), fraction de zone humide (FWL), fraction d'eau pure (FWP), fraction d'eau salée (FWS), fraction de sol nu ou aride (FEB) et la fraction de zone urbaine (FEU). Une part de diagramme détachée représente une influence négative et une part attachée une influence positive.

sable sur l'erreur SMOS est un résultat plutôt inattendu. Cette influence positive du sable est également retrouvée en Europe et en Amérique du Sud.

En Amérique du Sud, le pourcentage de sable et la fraction de forêt ont quasiment la même proportion d'influence avec 39% et 43%. Ainsi, la texture du sol joue un rôle majeur dans l'erreur SMOS sur ce continent. La texture du sol a également une grande influence en Europe (46%), en Afrique du Sud (46%), en Asie de l'Est (39%) et en Australie (54%). Cependant, cette influence de la texture n'est pas toujours représentée de la même manière : par l'argile avec une influence positive en Europe (43%) et Asie de l'Est (30%), avec une influence négative en Afrique du Sud (38%). Le sable a une influence négative en Australie (53%) et positive en Amérique du Sud (39%). A l'échelle globale, l'influence de la texture du sol est représentée par le sable négativement (22%).

Les RFI influencent l'erreur SMOS plus significativement en Asie centrale (47%), en Europe (24%), en Afrique du Nord (18%) et en Asie de l'Est (16%). Pour l'Asie de l'Est, ces statistiques auraient pu être bien plus élevées si on avait également pris en compte le nombre de points tellement affectés par les RFI qu'aucune valeur d'humidité n'a pu être retrouvée. Sur les trois autres continents cités, les RFI ne sont clairement pas négligeables, plus spécialement au centre de l'Asie où les interférences sont le premier facteur d'influence.

L'influence de la forêt en Australie peut paraître surprenant sur un continent qui ne possède pas de grande forêt dense comme l'Afrique. Cependant, les erreurs SMOS sont très faibles en Australie sauf près des côtes (fig. 4.7) où se situent les forêts les plus denses.

4.3.3 Classification des erreurs

Après avoir déterminé les principales sources des erreurs variables de SMOS avec une régression linéaire multiple, une classification de ces erreurs permet de prédire une erreur moyenne en fonction d'un ensemble de valeurs des différents paramètres. Ainsi il est possible de mettre en évidence des ensembles de valeurs qui conduisent à de très fortes ou de très faibles erreurs. La méthode utilisée dans cette partie est le CART (*Classification And Regression Tree*, Breiman et al. (1984)). La première étape de cette méthode est de créer des classes dont les valeurs des erreurs soient les moins dispersées. La seconde étape consiste de façon graphique à construire un arbre de décision (ou de régression) qui explique comment ces classes ont été construites.

4.3.3.1 Principe et théorie

Le principal objectif de la classification est de résumer et de prédire une variable par un ensemble de paramètres explicatifs. La classification se réalise récursivement en examinant chaque paramètre et chaque valeur seuil possible afin de créer les classes les plus homogènes possibles (les valeurs contenues dans ces classes sont donc de moins en moins dispersées). Dans notre cas, on essaie de créer des groupes rassemblant des valeurs des paramètres explicatifs (texture du sol, fraction de forêt, interférences,...) qui correspondent à un niveau équivalent d'erreur SMOS.

Soient x_i un paramètre explicatif et s_j une valeur de ce paramètre, alors le découpage $x_i < s_j$ et $x_i \geq s_j$ divise le jeu de données en deux ensembles disjoints que l'on peut appeler partitions ou classes. Les partitions continuent d'être créées jusqu'à ce qu'un certain niveau d'homogénéité dans toutes les classes créées soit atteint.

L'homogénéité à l'intérieur d'une classe est représentée par l'index de Gini I (Gini, 1912) qui mesure la dispersion des données contenues dans une classe. Si on dénote les classes par k ($k = 1, \dots, C$, où C est le nombre total de classes), alors l'index de Gini de la partition A est

défini par :

$$I(A) = 1 - \sum_{k=1}^C p_k^2 \quad (4.11)$$

où p_k est la fraction des observations qui appartiennent à la classe k à l'intérieur de la partition A . $I = 0$ représente la parfaite homogénéité où seulement une classe est présente dans la partition alors que $I = (C - 1)/C$ représente l'hétérogénéité complète où toutes les classes apparaissent équitablement dans la partition.

Si la somme des indices de Gini des deux possibles partitions est inférieure à l'indice de la partition à découper, c'est que l'homogénéité a été améliorée et on accepte le découpage. Afin d'éviter de trop partitionner et finalement de ne capturer que du bruit (sur-apprentissage), l'arbre de découpage ainsi formé est élagué. Le niveau d'élagage est contrôlé par le paramètre de complexité (cp). Tout découpage qui n'améliore pas le taux de variance expliquée d'un facteur cp n'est pas réalisé. Ce taux de variance expliquée est représenté par le coefficient de détermination qui est le carré du coefficient de corrélation (R^2). Ainsi, du temps de calcul est économisé puisque les découpages qui n'en valent pas la peine ne sont pas effectués. Dans cette étude, le coefficient de complexité cp est fixé arbitrairement à 0.01 aussi bien à l'échelle globale que continentale pour que les résultats restent comparables.

4.3.3.2 Résultats

4.3.3.2.1 A l'échelle globale. Un arbre de régression a été généré pour la Terre entière et se termine avec 6 feuilles (fig. 4.15). Ces feuilles représentent la fin de plusieurs décisions menant ainsi à une valeur moyenne de l'erreur SMOS. Cinq décisions ont été opérées et ont été considérées comme les plus discriminantes pour diviser le jeu de données global par l'algorithme CART. Les décisions menant à l'erreur SMOS la plus faible ($0.031 \text{ m}^3/\text{m}^3$) ont été : pas de présence de forêt dans le champ de vue du radiomètre ni d'eau libre (FFO = 0 et FWP = 0). Ces deux paramètres accompagnés de leur valeur seuil ont différencié au mieux le jeu de données. Au contraire, les décisions menant à l'erreur SMOS la plus forte ont été : FFO > 0 et RFI \geq 0.07. Si une forêt est détectée dans le champ de vue du radiomètre et que la probabilité d'interférence est supérieure à 7% en moyenne alors l'erreur moyenne SMOS est de $0.070 \text{ m}^3/\text{m}^3$.

Le premier découpage est extrêmement discriminant en terme de valeur moyenne de l'erreur puisque s'il n'y a pas de forêt, l'erreur moyenne est de $0.037 \text{ m}^3/\text{m}^3$ alors que s'il y a de la forêt l'erreur moyenne s'élève à $0.058 \text{ m}^3/\text{m}^3$. La décision *vraie* concerne un tiers de points alors que la décision *faux* rassemble les deux tiers des points. En présence de forêt, la deuxième décision porte sur la présence d'interférence. Une valeur seuil de 7% correspond à une probabilité d'occurrence moyenne sur 2010. Si un point a été affecté par les interférences moins de 7% dans ses observations, la décision suivante concerne encore la présence de forêt mais cette fois ci avec une valeur seuil plus élevée de 0.15, i.e., 15% du champ de vue du radiomètre est couvert de forêt. Si cette condition est *vraie*, alors une dernière décision concernant la présence ou non de zone humide se présente. Cette dernière décision n'est cependant pas négligeable puisque si elle est vraie, l'erreur moyenne est de $0.048 \text{ m}^3/\text{m}^3$ contre $0.066 \text{ m}^3/\text{m}^3$ si elle est fausse.

Les deux analyses (ANOVA et CART) ont donné des résultats complémentaires. L'ANOVA a calculé la part de la variance de l'erreur SMOS qui est expliquée par chaque paramètre. Le CART a déterminé par quelle variable avec quelle valeur seuil, le jeu de données des erreurs SMOS pouvait être découpé afin que chaque sous-ensemble puisse être le plus homogène possible et qu'il soit explicable différemment. Cela a créé une liste de décisions et l'erreur moyenne SMOS dépend de ces décisions. Même si certaines variables explicatives n'ont pas un impact important dans l'ANOVA, elles peuvent avoir une très grande influence sur le découpage du jeu de données

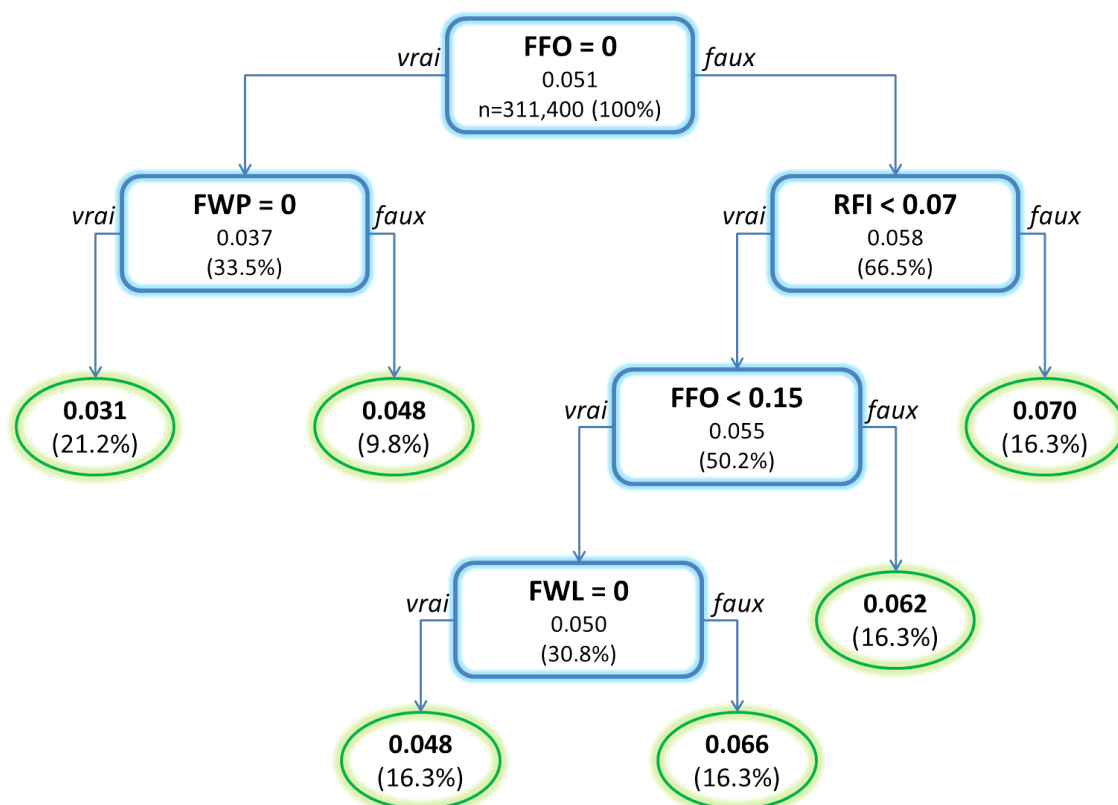


FIGURE 4.15 – Arbre de régression de l'erreur SMOS à l'échelle globale. Dans les rectangles sont inscrites les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre et/ou le pourcentage de points qui sont concernés. Si la condition est remplie, alors la branche *vrai* de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.

des erreurs SMOS. Par exemple, à l'échelle globale, les RFI n'expliquent que 1.7% de la variance de l'erreur SMOS d'après l'ANOVA alors que c'est le deuxième plus important paramètre dans l'arbre de décision qui mène à l'erreur la plus élevée d'après le CART.

4.3.3.2.2 A l'échelle continentale. Des arbres de régression comme celui de la figure 4.15 ont été générés pour chaque continent (annexe B). Les décisions menant aux valeurs moyennes les plus faibles et les plus élevées de l'erreur SMOS sont résumées dans le tableau 4.3. Le CART est stoppé de la même manière que précédemment à l'aide du paramètre de complexité qui est toujours fixé à 0.01. La fraction de forêt (FFO) est impliquée dans toutes les décisions continentales et cette fraction est dans la plupart des cas la première variable à être utilisée pour diviser le jeu de données.

L'Afrique du Nord, l'Afrique du Sud et l'Australie sont les trois continents où la dernière branche a mené aux erreurs moyennes SMOS les plus faibles. En Afrique du Nord, plus de 50% des points vérifient les conditions $\text{FFO} = 0$, $\text{FEU} = 0$ et $\text{FWP} = 0$ et obtiennent ainsi une erreur SMOS de $0.026 \text{ m}^3/\text{m}^3$ en moyenne. En Afrique du Sud, 17% des points vérifient la simple décision $\text{FFO} = 0$ pour une erreur moyenne de $0.031 \text{ m}^3/\text{m}^3$. Enfin en Australie, 40% des points satisfont les conditions $\text{FFO} = 0$ et $\% \text{argile} < 19.1$ pour une erreur moyenne de $0.033 \text{ m}^3/\text{m}^3$.

L'Asie de l'Est et l'Asie centrale sont les deux régions que le CART a identifié comme

Tableau 4.3 – Valeurs des paramètres qui mènent à l’erreur SMOS la plus faible et la plus élevée pendant le processus de classification. Les paramètres sont indiqués dans l’ordre d’apparition dans l’arbre de régression et les erreurs moyennes SMOS sont entre parenthèses.

	Erreur SMOS la plus faible	% points	Erreur SMOS la plus élevée	% points
Global	FFO= 0, FWP= 0 (0.031)	21.8	FFO> 0, RFI \geq 0.07 (0.070)	16.3
Am. du Nord	FFO= 0, FWP= 0 (0.037)	12.1	FFO > 0, %argile \geq 20.4 (0.067)	17.2
Am. du Sud	FFO= 0, RFI< 6.10^{-3} , %sable \geq 40, FEU= 0 (0.043)	23.9	FFO> 0, %sable< 44, FWP> 0 (0.070)	11.8
Europe	FFO= 0, RFI< 0.16, FWL= 0, FEU= 0 (0.038)	8.9	FFO> 0, RFI \geq 0.06 (0.068)	56.3
Af. du Nord	FFO= 0, FEU= 0, FWP= 0 (0.026)	50.6	FFO> 0 (0.064)	22.1
Af. du Sud	FFO= 0 (0.031)	16.7	FFO \geq 0.32, %argile< 17.4 (0.068)	18.7
Asie centrale	RFI< 0.11, FFO= 0 (0.038)	24.0	RFI \geq 0.11, FFO> 0 (0.076)	30.6
Asie de l’Est	RFI< 0.20, FFO= 0 FWL= 0 (0.049)	10.0	RFI \geq 0.20, %argile \geq 22.2 (0.080)	20.6
Australie	FFO= 0, %argile< 19.1 (0.033)	39.7	FFO \geq 0.24 (0.061)	8.6

étant celles menant aux erreurs les plus élevées en moyenne. Sur l’Asie de l’Est, les conditions $\text{RFI} \geq 0.20$ et $\% \text{argile} \geq 22.2$ sont remplies par 21% des points pour une erreur SMOS moyenne de $0.080 \text{ m}^3/\text{m}^3$. En Asie centrale, 31% des points vérifient les conditions $\text{RFI} \geq 0.11$ et $\text{FFO} > 0$ pour une erreur moyenne de $0.76 \text{ m}^3/\text{m}^3$.

On peut remarquer que la classification pour l’Amérique du Sud est cohérente avec les résultats de l’ANOVA sur ce continent. Sur cette région, le sable avait une influence positive, il est donc normal de retrouver ici que l’erreur la plus basse est obtenue quand le sol en contient au moins 40%.

4.4 Conclusions et perspectives

Le produit d’humidité SMOS est disponible depuis le 12 Janvier 2010. Dans cette étude, la première année de données a été utilisée pour évaluer la structure de l’erreur variable de SMOS à l’échelle globale. Afin d’y parvenir, la méthode de la triple collocation a été appliquée à SMOS et d’autres jeux de données d’humidité qui couvraient l’ensemble du globe également : AMSR-E(VUA), AMSR-E(NSIDC), ASCAT et le modèle ECMWF. Les cartes d’erreur ont montré que SMOS donne de meilleurs résultats (i.e., des erreurs variables plus faibles) que les autres jeux de données sur une partie de l’Amérique du Nord, en Australie et dans l’Ouest de l’Asie.

La méthode de triple collocation ne tient pas compte de l’erreur de biais des produits testés,

ce qui avantage les produits d'humidité dérivés à partir d'algorithme détectant les changements de niveaux d'humidité. D'après les résultats du chapitre précédent, l'essentiel de l'erreur SMOS se situe dans la variance alors que les autres produits présentent un grand biais. Or la triple collocation ne va traiter que la partie variables des produits d'humidité (utilisation des anomalies), ce qui explique des résultats en défaveur de SMOS.

Bien que les erreurs issues de la triple collocation appliquée à un triplet ne peuvent pas être comparées à des erreurs issues d'un autre triplet, la structure spatiale de l'erreur SMOS à l'échelle globale est la même pour tous les triplets testés : erreurs élevées en Asie de l'Est et du Sud, en Europe et sur les zones à forte végétation (Amazonie, Afrique centrale, extrême Nord de l'Amérique) et des erreurs faibles en Afrique du Nord, la pointe Sud de l'Afrique, dans l'ouest de l'Asie, en Australie et sur une partie de l'Amérique du Nord. En obtenant des cartes d'erreur similaires, on peut supposer que la triple collocation est une méthode robuste au choix du triplet.

Le deuxième objectif de ce chapitre était de relier les valeurs et la structure des erreurs variables de SMOS à des paramètres physiques comme la texture du sol (pourcentages de sable et d'argile), la probabilité d'occurrence d'interférences (RFI) et l'occupation du sol telle que défini par ECOCLIMAP (fractions de forêt, de zone humide, d'eau libre, d'eau salée, de sol nu ou aride, de glace et de zone urbaine vues par le radiomètre). Une régression linéaire multiple visant à expliquer l'erreur variable de SMOS linéairement par ces paramètres a été réalisée sur chaque continent ainsi qu'à l'échelle globale. Une analyse de la variance (ANOVA) a déterminé par la suite la proportion de variance de l'erreur qu'explique chaque paramètre.

A l'échelle globale, la fraction de forêt explique la plupart de la variance de l'erreur SMOS (57%) suivi de la texture du sol avec le pourcentage de sable (22%). Plus le radiomètre voit de la forêt ou du sol contenant beaucoup de sable, plus l'erreur SMOS sur l'humidité est grande. Ces proportions varient grandement suivant les continents. En Europe, la proportion de variance expliquée par la texture est de 46% suivie par les interférences avec 24% alors qu'en Amérique du Nord, la fraction de forêt expliquait 67% de la variance totale. Il est important de noter aussi qu'une grande proportion correspond également au besoin d'une grande précision. Prenons l'exemple de l'Australie où le pourcentage de sable explique 52% de la variance de l'erreur SMOS, si ce paramètre est mal évalué au départ et n'est pas précis, cela aura un impact important sur l'erreur SMOS puisque plus de la moitié de la variance de cette erreur est expliquée par la variance de la texture du sol. Par conséquent, des cartes très précises de la texture et de l'occupation du sol sont nécessaires à l'échelle globale.

La présence d'interférence est un grand problème qu'il ne faut pas sous-estimer car elle a un sérieux impact sur les valeurs d'humidité retrouvées par SMOS. Même si la proportion expliquée par les RFI de la variance de l'erreur est faible à l'échelle globale, son influence est très grande en Asie, en Europe et en Afrique du Nord. De plus, les régions du globe les plus affectées par ces RFI n'ont pas d'humidité retrouvée car les observations sont bien trop polluées pour pouvoir retirer une quelconque information biophysique. Si le nombre de points où l'humidité ne peut pas être retrouvée à cause des interférences avait été pris en compte, l'influence des RFI aurait été bien plus grande.

Afin d'identifier le groupe de paramètres qui mène aux erreurs les plus faibles et aux erreurs les plus élevées, une classification et un arbre de régression (CART) ont été réalisés sur les résultats de chaque continent ainsi qu'à l'échelle globale. A l'échelle globale, et pour la plupart de continents, la fraction de forêt vue par le radiomètre est le premier paramètre à être utilisé pour diviser le jeu de données. Sauf en Amérique du Sud, la valeur seuil de la fraction de forêt est très basse (max. 0.07) donc même une très petite fraction de forêt peut mener à une valeur très élevée de l'erreur SMOS. Dans l'ouest de l'Asie et en Asie de l'Est, les RFI sont le premier paramètre qui permet de faire la première division. Ces deux régions sont extrêmement affectées

par les interférences.

Dans cette étude, les paramètres qui évoluent avec le temps n'ont pas été pris en compte comme la présence de gel ou de neige et peuvent jouer un rôle majeur dans l'explication de l'erreur SMOS, essentiellement sur les régions aux hautes latitudes. Puisque l'algorithme SMOS utilise les valeurs ECMWF pour estimer les contributions des différentes classes de couverture du sol, il est possible que ECMWF influence les humidités retrouvées par SMOS et que les anomalies de ces deux jeux de données soient corrélées. Ceci peut poser un problème lors de la triple collocation car si les deux anomalies sont corrélées il faut ajouter un terme lors du calcul de la covariance qui n'est pas facile à estimer. Cependant, les résultats globaux et les structures spatiales de l'erreur sont très similaires pour les trois triplets donc l'analyse finale de l'erreur SMOS reste inchangée.

Un point qui n'a pas été couvert par cette étude est l'influence de l'uniformité ou de la non-uniformité de la surface observée sur l'humidité retrouvée. En effet, une scène contenant 20% de forêt et 80% de végétation basse peut être perçue différemment si la forêt est regroupée ou si elle est dispersée dans le champ de vue de l'instrument. Il est cependant attendu que si la scène est très hétérogène avec plusieurs classes représentées, les contributions de chaque classe doivent être estimées et cela introduit d'autant plus d'erreur et d'incertitude dans l'humidité retrouvée.

A partir des résultats de cette étude, plusieurs perspectives peuvent être envisagées concernant l'amélioration de l'algorithme SMOS. La variation de la fraction de forêt explique plus de la moitié de la variance de l'erreur à l'échelle globale et deux conséquences sont à prévoir : l'utilisation d'une carte plus précise de l'occupation du sol (pas nécessairement au point de vue de la résolution mais plutôt au niveau de la précision de la mesure) et une révision de la paramétrisation du modèle de la forêt qui permet de modéliser sa contribution au signal. Le deuxième paramètre le plus important est la texture du sol et une amélioration future pour les humidités SMOS serait d'utiliser une carte plus précise en terme de résolution plutôt que d'utiliser une carte grossière comme c'est le cas aujourd'hui dans l'algorithme SMOS. Cependant, la texture ne joue pas un rôle uniforme dans toutes les régions du monde. Le sable a une influence négative en Afrique, en Asie et en Australie alors qu'il a une influence positive en Amérique et en Europe. Pour les régions où le sable joue un rôle négatif, l'utilisation du modèle de constante diélectrique de Mironov (Mironov & Fomin, 2009) devrait améliorer les résultats. SMOS utilise actuellement le modèle de Dobson (Dobson et al., 1985) mais une nouvelle version du processeur (v5.6) est parue avec l'implémentation du modèle de Mironov par défaut, le modèle de Dobson devenant une option. Les résultats confirment le besoin d'avoir le choix entre ces deux modèles dépendant du type de sol. De plus, un retraitement des températures de brillance acquises par SMOS avec une calibration unique sur 2 ans a été réalisée, ce qui devrait améliorer la stabilité des valeurs d'humidité tout au long de ces deux années.

L'étude menée dans ce chapitre s'est concrétisée sous forme d'un article soumis au journal *Remote Sensing of Environment* le 7 Mars 2012 appelé *Spatial distribution and possible source of SMOS errors at the global scale* et écrit par Delphine J. Leroux, Yann H. Kerr, Philippe Richaume et Rémy Fieuzal (CESBIO, Toulouse). J'ai également pu présenter ces travaux sous la forme d'un oral aux *IGARSS (International Geoscience And Remote Sensing Symposium, Vancouver, Canada)* en 2011 et d'un poster au *SMOS Workshop (Arles, France)* en 2011.

Chapitre 5

Construction d'un historique de l'humidité du sol à partir de SMOS

L'humidité des sols joue un rôle important dans le cycle hydrologique car elle contribue directement aux transferts d'eau entre le sol et l'atmosphère. L'impact du dérèglement climatique sur le cycle de l'eau est encore inconnu : ralentissement ou accélération. Ce type d'étude climatique requiert de longues séries temporelles stables qui permettent d'observer l'évolution d'une variable en particulier.

Depuis 1978, de nombreux satellites ont permis d'observer la Terre dans sa globalité et ainsi retrouver l'humidité à sa surface ces trente dernières années. Malgré la valeur inestimable de cet historique, cette longue série temporelle de l'humidité n'est pas utilisable à cause des nombreux sauts de valeurs lors des changements de capteurs. Lorsque deux satellites ont des durées de vie qui se chevauchent, il est alors possible d'établir un lien entre leurs produits d'humidité.

Dans ce chapitre, deux méthodes statistiques sont présentées afin de relier les humidités SMOS (*Soil Moisture and Ocean Salinity*) et AMSR-E (*Advanced Microwave Scanning Radiometer*) sur les quatre sites test aux États-Unis présentés dans le chapitre 3. Le CDF matching (*Cumulative Density Function*) et les copules sont deux méthodes basées sur la même théorie statistique mais le CDF matching suppose que les deux séries temporelles à homogénéiser sont ordonnées de la même façon alors que les copules ne posent pas cette hypothèse.

L'année 2010 est utilisée comme référence afin de calibrer les deux modèles statistiques. Une série temporelle homogène est ainsi générée pour chaque site test de 2003 à 2010 qui est finalement validée avec des mesures terrain d'humidité.

Le CDF matching et la méthode des copules gardent le même niveau de corrélation que les humidités originales d'AMSR-E mais diminuent l'erreur par rapport aux mesures terrain. La méthode des copules apporte cependant une diminution de 25% en moyenne en terme d'erreur par rapport au CDF matching sur la longue série temporelle de 2003 à 2010.



5.1 Introduction

5.1.1 Etat de l'art

L'humidité du sol joue un rôle majeur dans les transferts d'eau et d'énergie entre le sol et l'atmosphère. L'humidité est également une variable d'entrée pour les modèles de surface qui permettent de modéliser la fraction évaporative à la surface du sol et l'infiltration dans la zone racinaire. Pour une meilleure gestion de l'eau dans l'agriculture, l'information d'humidité est essentielle aussi bien à l'échelle locale que régionale. A l'échelle globale, l'humidité du sol est une information très utile pour la prévision météorologique (Drusch, 2007), l'étude du changement climatique (Douville & Chauvin, 2000) et l'observation et la prédiction d'événements extrêmes comme les inondations ou les sécheresses. La gestion de l'eau est une problématique très importante d'autant plus que la population ne cesse de croître (Prigent et al., 2012).

Grâce à leur couverture globale et leur haute répétitivité, les satellites sont d'une grande valeur et leurs informations sont de plus en plus utilisées. De longues séries temporelles basées sur ces informations satellite sont ainsi profitables pour l'analyse du climat en général. L'utilisation des observations acquises par les satellites de la génération précédente implique cependant la résolution de certains problèmes. Certains concepts ont été développés pour retrouver l'humidité à partir de fréquences plus élevées que celles de SMOS puisque cela représentait la seule option jusqu'à maintenant. Parmi les satellites précédents, on trouve SMMR (*Scanning Multichannel Microwave Radiometer*, 1978-1987, Njoku et al. (1980), Owe et al. (2001)), SSM/I (*Special Sensor Microwave/Imager*, 1987-, Barrett et al. (1988), Owe et al. (2001)), AMSR-E (*Advanced Microwave Scanning Radiometer-Earth observation system*, 2002-2011, Owe et al. (2001), Njoku et al. (2003), Kawanishi et al. (2003)), WindSat (2003-, Gaiser et al. (2004), Li et al. (2010b)) et ASCAT (*Advanced Scatterometer*, 2006-, Figa-Saldana et al. (2002), Naeimi et al. (2009)). Bien que leurs plus basses fréquences (5-20 GHz) ne soient pas les plus adaptées pour retrouver l'humidité à cause de leur grande sensibilité à la végétation et aux conditions atmosphériques, les informations contenues dans leurs acquisitions sont cruciales et représentent tout de même un historique de données de 1978 à nos jours (fig. 5.1).

Les applications environnementales comme l'assimilation de données ou l'étude du changement climatique ont besoin cependant d'historiques stables dans la durée. Or les produits d'humidité cités précédemment sont retrouvés à partir d'instruments différents avec des algorithmes différents, et cela résulte en une série temporelle de l'humidité non homogène et instable. La figure 5.2 représentant l'historique de l'humidité dans le désert du Namib en Afrique est un exemple (données SMMR, SSM/I, AMSR-E, ASCAT et SMOS). Cette non-homogénéité pourrait être interprétée comme un biais dans un processus d'assimilation de données. Pour éviter ce problème, ces différents produits ont besoin d'être modifiés pour corriger les biais existants et les variations d'amplitude entre les jeux de données.

Plusieurs études ont déjà été menées sur l'homogénéisation de séries temporelles. Vincent et al. (2002) ont développé une méthode pour harmoniser des séries temporelles de températures qui comportaient des sauts ou des variations trop brusques. La première étape était de déterminer si la série était homogène en comparant ses anomalies (par rapport à la moyenne temporelle) à celles d'un historique *référence*. L'identification des sauts et de leurs amplitudes a été réalisée en appliquant successivement un modèle linéaire avec différentes valeurs d'amplitude et le meilleur ajustement était celui qui obtenait la plus petite somme des erreurs au carré. Des séries temporelles homogènes de température et de précipitation ont été développées par Begert et al. (2005) en utilisant des méthodes statistiques pour détecter des potentielles in-homogénéités. Encore une fois, une série temporelle de référence était nécessaire pour détecter et calculer l'amplitude de ces sauts. Picard & Fily (2006) ont proposé une méthode pour simuler des séries temporelles homogènes de la fonte cumulée de la surface en Antarctique. L'utilisation d'obser-

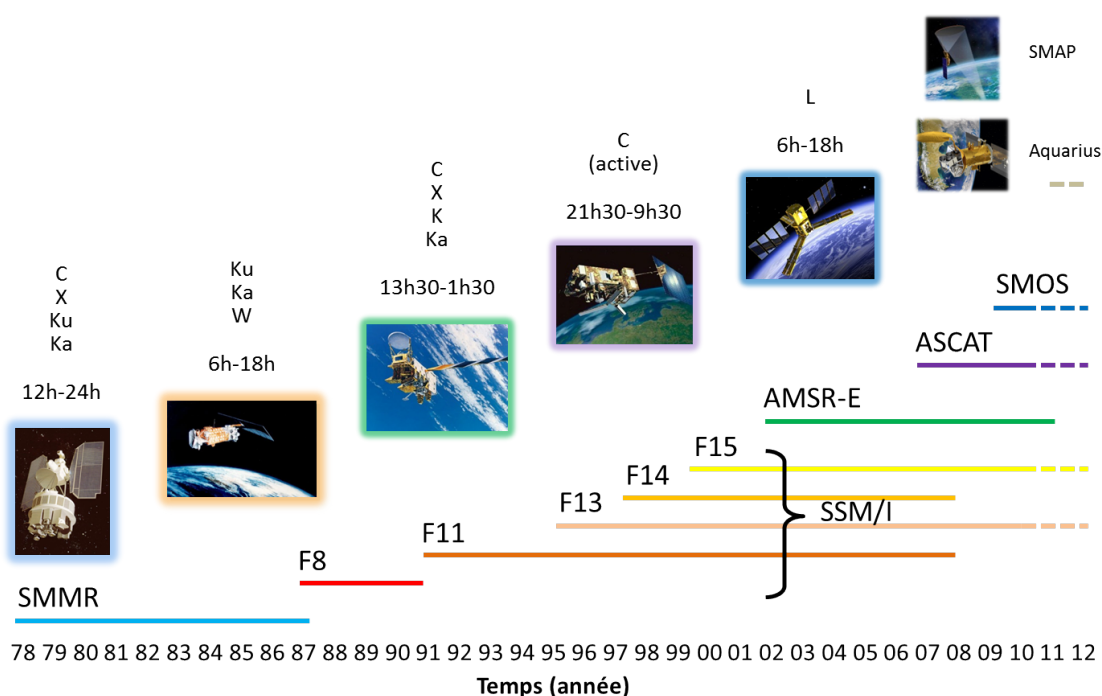


FIGURE 5.1 – Représentation temporelle de quelques satellites qui ont précédé ou qui vont suivre la mission SMOS pour retrouver l’humidité des sols à l’échelle globale : SMMR (*Scanning Multi-channel Microwave Radiometer*, 1978-1987), SSM/I (*Special Sensor Microwave/Imager*, 1987-, satellites F8 à F15), AMSR-E (*Advanced Microwave Scanning Radiometer-Earth observation system*, 2002-2011), ERS-ASCAT (*Advanced Scatterometer*, 2006-), SMOS (*Soil Moisture and Ocean Salinity*, 2009-), Aquarius (2011-), SMAP (*Soil Moisture Active Passive*, prévu en 2014). Les bandes de fréquence et les heures de passage sont également indiquées pour chaque mission spatiale.

vations provenant de différents satellites a été le plus grand défi dans cette étude. La correction de l’effet de l’heure d’observation a été effectuée en deux étapes. Une fonction sinusoïdale de période 24h a d’abord été ajustée aux observations puis une interpolation optimale pour raffiner le modèle de première hypothèse pour le forcer à être plus près des observations, quitte à se détacher du modèle sinusoïdal de départ, afin d’obtenir une incertitude faible autour des heures d’observation et une incertitude forte quand il n’y pas d’observation.

Faire correspondre les distributions statistiques (ou *CDF matching*) est une méthode déjà utilisée dans de nombreuses études pour construire des séries temporelles homogènes. Reichle & Koster (2004), Choi & Jacobs (2008) ont fait correspondre des humidités retrouvées à partir d’observations satellites avec des données de modèle, Li et al. (2010a) ont corrigé les biais des produits de précipitation et de température retrouvés à partir de différents modèles. Le *CDF matching* a également été utilisé comme étape préliminaire pour l’assimilation de données (Drusch et al., 2005) et pour produire des longues séries temporelles d’humidité (Liu et al., 2009, 2011).

Au cours des dernières années, une nouvelle méthode basée sur l’utilisation de fonctions statistiques appelées *copules* a été développée (Nelsen, 1998). La fonction copule sert à relier des distributions statistiques entre elles. Il est ainsi possible de dériver des distributions bi-variées en ne faisant aucune hypothèse concernant la nature des distributions : elles ne sont pas obligatoirement normales ou du même type et il n’est pas nécessaire que les variables soient

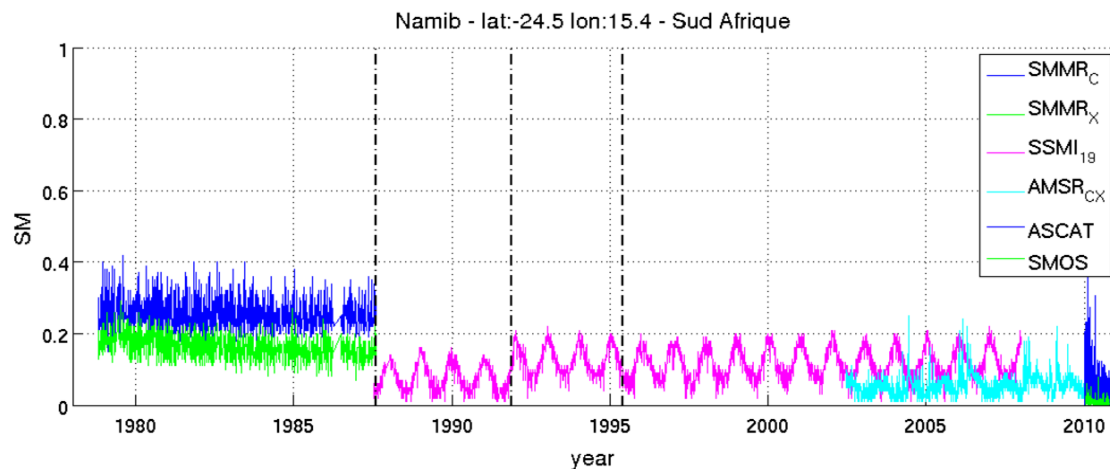


FIGURE 5.2 – Série temporelle de l'humidité des sols dans le désert du Namib en utilisant les observations des satellites SMMR, SSM/I, AMSR-E, ASCAT et SMOS (en m^3/m^3).

indépendantes. Le principal avantage de la méthode des copules est de pouvoir relier toute forme de distribution (Singh & Strupczewski, 2007). La première vulgarisation et démocratisation des fonctions copules a été écrite par Nelsen (1998). Il présente plusieurs méthodes pour la construction de ces fonctions et discute de leur rôle dans la modélisation de la dépendance entre plusieurs variables. Depuis, les copules ont été utilisées dans de nombreux domaines dont une majorité des applications dans le domaine de la finance (Frees & Valdez, 1998, Armstrong, Trivedi & Zimmer, 2005, Arnold, 2006, Hafner & Reznikova, 2010).

En hydrologie, plusieurs applications ont commencé à voir le jour. Genest & Favre (2007) ont listé les méthodes existantes pour détecter et évaluer la dépendance entre plusieurs jeux de données à travers les copules de manière analytique et graphique, et ont présenté plusieurs méthodes pour choisir la famille de copules qui s'adapte le mieux aux jeux de données utilisés ainsi que pour l'estimation des paramètres associés au choix de famille. Favre et al. (2004) ont appliqué la méthode des copules à des débits et des volumes d'eau mesurés sur deux bassins versants afin d'étudier leur dépendance, Salvadori & de Michele (2007) aux séries temporelles de tempêtes et de précipitations, Dupuis (2007) aux volumes et temps de parcours des flux lents de deux rivières, Zhang & Singh (2007) aux fréquences d'événements pluvieux, Serinaldi & Grimaldi (2007) aux fréquences d'inondation et Laux et al. (2011) aux données de précipitation. Gao et al. (2007) ont également utilisé les copules comme étape préliminaire dans un processus d'assimilation de données afin de normaliser les variables d'entrée.

Ce type d'analyse statistique (analyse statistique jointe) a déjà été utilisé lorsque les données d'humidité proviennent de différents systèmes d'observation, comme par exemple dans les travaux de Reichle & Koster (2004) où des humidités de surface retrouvées à partir d'observations AMSR-E sont comparées à des humidités à 10 cm retrouvées à partir d'un modèle de surface. Il existe plusieurs autres études utilisant les statistiques jointes sur des variables qui ne sont pas physiquement liées mais où leur relation statistique peut être utile (de Michele & Salvadori, 2003).

5.1.2 Objectif

Pour étudier le climat à travers les séries temporelles d'humidité du sol à l'échelle globale, il est nécessaire d'utiliser des données satellite ou des sorties de modèles météorologiques. Les valeurs d'humidité ont été retrouvées avec la méthode de Owe et al. (2001) à partir d'observa-

Tableau 5.1 – Coefficient de corrélation (R) entre les mesures terrain d’humidité à 1:30h et 6:00h sur les quatre bassins versants test en 2010 (Walnut Gulch, Little Washita, Little River et Reynolds Creek) avec son intervalle de confiance à 95%. N est le nombre de dates communes disponibles en 2010.

Walnut Gulch		Little Washita		Little River		Reynolds Creek	
R	N	R	N	R	N	R	N
0.96 (0.95-0.97)	365	0.97 (0.96-0.98)	365	0.95 (0.94-0.96)	365	0.99 (0.99-0.99)	328

tions provenant de différents satellites et bien que leurs fréquences d’utilisations ne soient pas toutes adéquates pour retrouver l’humidité du sol, cette série temporelle n’en reste pas moins intéressante à étudier. Cependant on remarque des sauts et des changements de dynamique dans la série temporelle lors des changements de capteur essentiellement. Le plus grand saut se trouve au passage de SMMR et de SSM/I (1987, fig. 5.2) alors que des dynamiques inversées sont visibles sur la période de recouvrement entre SSM/I et AMSR-E : SSM/I observe une humidité plus élevée en hiver alors qu’AMSR-E observe le contraire (2002-2008, fig. 5.2). A travers cet exemple, on peut comprendre l’utilité et la nécessité de reconstruire des séries temporelles homogènes de l’humidité en s’appuyant sur les historiques déjà existants.

Le but de ce chapitre est d’estimer pour toute la période d’AMSR-E (2003-2010) ce qu’aurait mesuré SMOS dans le but de construire une longue série temporelle d’humidité homogène de 2003 à nos jours. De cette façon, cette longue série temporelle serait utilisable dans des applications telles que le suivi des sécheresses qui est basé sur les centiles statistiques. Pour cela, la première étape est de comprendre le lien et l’inter-dépendance entre les données SMOS et AMSR-E sur leur période commune. L’année 2010 servira de référence et d’apprentissage pour la reconstruction des années précédentes.

Cependant, comparer et faire correspondre les observations d’AMSR-E acquises à 1:30h en bande C/X avec celles de SMOS acquises à 6:00h en bande L soulèvent quelques interrogations sur la faisabilité : (1) les temps de passage sont différents et des événements pluvieux peuvent se produire entre les deux acquisitions, (2) les fréquences d’observations sont différentes donc les profondeurs de pénétration du signal ne sont pas équivalentes. Les prochains paragraphes tentent de répondre à ces remarques.

5.1.3 Différences de temps et de fréquence d’acquisition

L’impact statistique des pluies qui pourraient intervenir entre 1:30h et 6:00h serait une diminution de la corrélation entre les deux variables d’humidité que l’on souhaite relier, i.e. AMSR-E et SMOS. Si cette corrélation est suffisamment élevée, une relation peut être établie pour estimer une valeur SMOS à partir d’une observation AMSR-E. Les mesures d’humidité réalisées sur le terrain à 1:30h et 6:00h des quatre bassins versants test ont été statistiquement analysées afin de déterminer si les pluies qui auraient pu intervenir entre ces heures ont eu un impact sur le coefficient de corrélation (tab. 5.1). Les données de précipitation sur ces sites test n’étant pas disponibles, les mesures d’humidité ont été utilisées. Tous les coefficients de corrélation se situent entre 0.95 et 0.99 et on peut donc supposer dans ce cas que les précipitations n’affecteront pas significativement l’analyse.

L’impact de la différence de fréquence d’observation entre SMOS et AMSR-E est faible dans la majorité des situations. La fréquence AMSR-E la plus élevée utilisée dans ce chapitre (10.7 GHz) se traduit par une profondeur d’émission plus superficielle que pour les observations

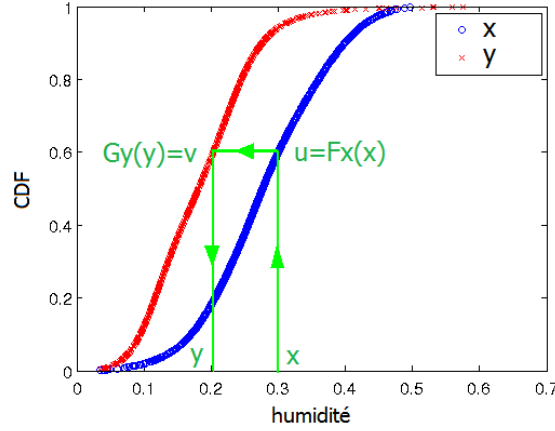


FIGURE 5.3 – Principe du CDF matching où l'égalité des probabilités est imposée. Pour une valeur x donnée, trouver la valeur y telles que les fonctions de répartitions soient égales : $G_Y(y) = F_X(x)$ ou $y = G_Y^{-1}(F_X(x))$.

SMOS (1.4 GHz), mais quand bien même les valeurs d'humidité sont différentes, leurs valeurs relatives sont similaires, soit humide ou sec. La corrélation entre deux variables n'est basée que sur leurs valeurs relatives (avec leur série temporelle individuelle) et pas sur leurs valeurs absolues. Dans le cas des copules, la corrélation est basée sur le rang occupé par chaque valeur donc cette différence de fréquence d'observations n'aura que très peu d'impact.

Si les deux profondeurs d'émission devaient être accordées, il faudrait mettre en place une relation qui permettrait de relier les mesures d'humidité à 5 cm à celles effectuée à 1 cm. Un tel modèle introduirait cependant une incertitude significative très difficile à estimer compte tenu de la variabilité (spatiale et en profondeur) des propriétés du sol dont les propriétés de séchage et de ressuyage sont différentes. Si les données SMOS (ou ASMR-E) devaient être ajustées à la profondeur d'émission d'ASMR-E (ou SMOS) par assimilation de données dans un modèle de surface par exemple, alors le jeu de données entier devrait être ajusté en insérant une incertitude due à cette étape supplémentaire. Il y aura donc une discordance entre le passé et le futur, peu importe quelle variable est adaptée à l'autre. La seule approche cohérente dans le temps est de traiter les données d'origine.

L'assimilation de données pourrait néanmoins résoudre ces différences de temps de passage et de fréquence mais cela impliquerait d'ajouter de l'incertitude impossible à estimer précisément. Cela impliquerait également l'utilisation de jeux de données de précipitations ainsi que des données météorologiques qui ajouteraient encore de l'incertitude au traitement. L'assimilation de données peut être vue comme une approche concurrente aux méthodes statistiques pour la construction d'un historique homogène de l'humidité des sols.

5.2 Méthodes statistiques pour créer une série temporelle homogène

Deux méthodes statistiques sont présentées et utilisées pour créer une série temporelle homogène de l'humidité des sols. Le CDF (*Cumulative Density Function*) matching a déjà été utilisé à plusieurs reprises dans d'autres travaux pour fusionner deux séries temporelles (Reichle & Koster, 2004, Liu et al., 2009, 2011) alors que la méthode des copules n'a été que très récemment utilisée pour des études de variables environnementales.

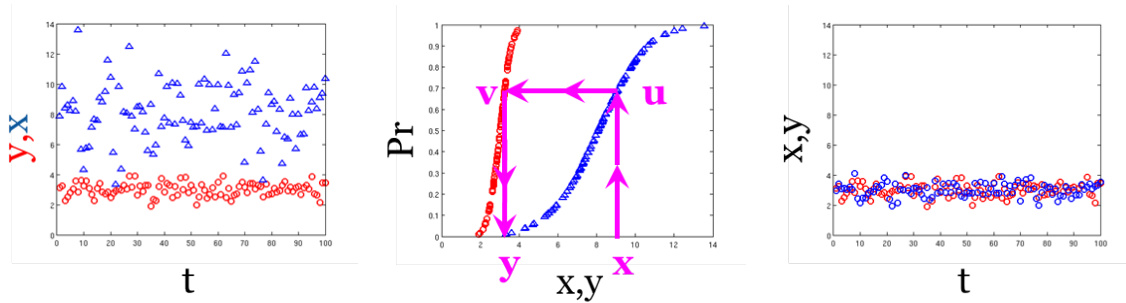


FIGURE 5.4 – Exemple d’application du CDF (*Cumulative Density Function*) matching. La série temporelle bleue est ramenée au niveau de la série rouge en faisant correspondre leur distributions statistiques.

5.2.1 CDF matching

La fonction de répartition (ou *cumulative density function* en anglais, CDF) est la probabilité qu’une variable aléatoire X prenne une valeur inférieure ou égale à une valeur donnée x :

$$F_X(x) = \Pr[X \leq x] \quad (5.1)$$

où F_X est la fonction de répartition de la variable aléatoire X . Si on considère deux séries temporelles, le CDF matching consiste à faire correspondre les fonctions de répartition des deux variables aléatoires en imposant l’égalité de leurs probabilités (fig. 5.3). La méthodologie suivante a été appliquée :

1. calculer les fonctions de répartition des deux ensembles de données X et Y : F_X et G_Y ,
2. pour chaque valeur x de l’ensemble X , trouver la valeur y dans Y telle que $G_Y(y) = F_X(x)$.

Concrètement, la figure 5.4 représente un exemple très simple où le but est de ramener la série temporelle bleue au niveau de la série rouge. Dans un premier temps, on calcule les fonctions de répartition de chaque série puis, pour chaque point bleu x , on trouve la probabilité u qui lui correspond. Pour trouver son équivalent dans la série rouge, on part graphiquement à l’horizontal depuis la valeur u jusqu’à arriver à une valeur rouge v (ce qui correspond à $u = v$, fig. 5.4 du milieu). On répète cette étape pour chaque point bleu et on peut ainsi reconstituer la série temporelle bleue au niveau de la série rouge (fig. 5.4 de droite).

Cependant, l’hypothèse que les probabilités $F_X(x)$ et $G_Y(y)$ peuvent être mises à égalité n’est jamais vérifiée puisque cela correspondrait à deux séries de valeurs *rangées* dans le même ordre, c’est à dire que l’ordre chronologique de chaque jeu de données serait identique : la valeur la plus faible de X aurait été observée à la même date que la plus faible valeur de Y , les 2ème valeurs les plus faibles de X et de Y auraient été également observées à la même date et ainsi de suite. Une même organisation des données (ou une même chronologie) justifierait de poser $u = v$. Or la plupart du temps, les deux ordres chronologiques de X et de Y ne sont pas identiques et lorsque l’on compare les deux fonctions de répartition, elles sont très dispersées comme sur la figure 5.5. La méthode des copules permet de modéliser cette dispersion qui contient de l’information relatant la dépendance entre ces deux variables. Dans le reste de l’étude, on définit les variables U et V pour représenter $F_X(x)$ et $G_Y(y)$. U et V sont les jeux ou ensembles de données alors que u et v sont des valeurs de ces ensembles.

5.2.2 Méthode des copules

La théorie des copules est un outil très utile et puissant pour modéliser la structure de dépendance entre deux ensembles de variables aléatoires. Comme le CDF matching, les co-

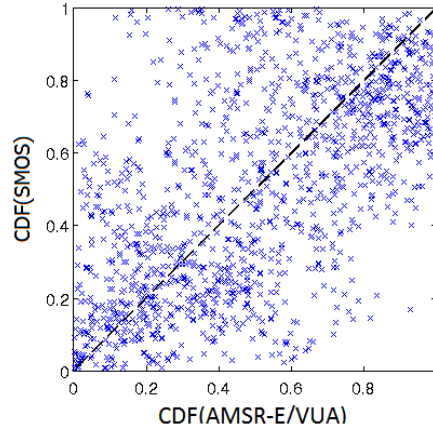


FIGURE 5.5 – Exemple de comparaison des fonctions de répartition (CDF) de SMOS et d'AMSR-E/VUA où les probabilités sont très dispersées et où l'hypothèse de les égaliser avec un CDF matching ferait perdre de l'information.

pules séparent le comportement marginal ou extrême des variables aléatoires de la structure de dépendance en utilisant les fonctions de répartition. En revanche, au lieu d'égaliser les probabilités u et v en faisant correspondre exactement les fonctions de répartition, les variables U et V sont comparées et analysées. La fonction copule sert de liaison entre les deux variables, d'où son nom.

5.2.2.1 Théorie générale

Une copule est une fonction qui génère une fonction de répartition ou distribution multivariée à partir de plusieurs autres fonctions de répartition. Soient deux variables aléatoires X et Y avec leur fonction de répartition respective F_X et G_Y , alors le théorème de Sklar formule :

$$H_{XY}(x, y) = C_{XY}(F_X(x), G_Y(y)) \quad (5.2)$$

$$= C_{XY}(u, v) \quad (5.3)$$

$$= \Pr[X \leq x, Y \leq y] \quad (5.4)$$

où H_{XY} est la fonction de répartition jointe de X et Y et C_{XY} la fonction de copule associée. Il est alors possible d'en déduire les distributions conditionnelles $H_{XY}(y | x)$, i.e. la fonction de répartition jointe sachant x , en utilisant les dérivées partielles. Soit $u = F_X(x)$ et $v = G_Y(y)$ alors $H_{XY}(y | x)$ ou $C_{V|U}$ sont définies par :

$$C_{V|U} = \frac{\partial C(u, v)}{\partial u} \quad (5.5)$$

Schweizer & Wolff (1981) ont établi que la fonction copule permet de rendre compte n'importe quelle forme de dépendance entre deux variables. Ils réussissent à démontrer qu'aucune transformation des variables X et Y n'affectera leurs variables associées U et V . Ainsi la façon dont les variables X et Y évoluent ensemble est obligatoirement capturée par la copule, quelle que soit l'échelle à laquelle chaque variable est mesurée.

5.2.2.2 Familles de copules

Il existe de nombreuses familles de copules qui ont chacune des propriétés différentes. La forme de la dispersion entre U et V va déterminer le choix de la famille qu'il faudra utiliser. La

largeur de la dispersion est ensuite contrôlée par un seul paramètre θ . La plupart des définitions qui vont suivre sont basées sur les travaux de (Nelsen, 1998).

La famille de copules qui permet de modéliser l'indépendance entre deux variables est le produit :

$$C(u, v) = u \cdot v \quad (5.6)$$

Une copule Archimédienne est définie par :

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)) \quad (5.7)$$

où ϕ est la fonction génératrice qui va de $[0, 1]$ à $(0, \infty)$ et satisfait trois conditions : $\phi(1) = 0$, ϕ strictement décroissante et ϕ convexe.

Les copules elliptiques ont des distributions avec des contours elliptiques. L'avantage principal des distributions elliptiques est que le niveau de corrélation entre les variables peut être spécifié. En revanche, les copules elliptiques sont contraintes d'avoir une symétrie radiale.

Dans cette étude, 9 familles de copules ont été testées : le produit ou indépendance, Clayton, Frank, Gumbel, Farlie-Gumbel-Moregenstern (FGM), Ali-Mikhail-Haq (AMH), Arch12 (la 12ème copule présentée dans (Nelsen, 1998)), Arch 14 (la 14ème copule présentée dans (Nelsen, 1998)) et la copule Gaussienne. Ces copules ont été choisies par rapport à ce qui existe déjà dans la littérature. Nelsen (1998) en a listé beaucoup plus et de nouvelles familles ont été créées depuis. Cependant un tri a été effectué de telle façon que chaque copule testée soit suffisamment différente des autres pour qu'elle soit reconnaissable. Ces 9 copules sont décrites dans le tableau 5.2 et la figure 5.6 et ont chacune leurs propres caractéristiques dont en voici quelques unes :

- Clayton : dépendance très forte à gauche et relativement faible à droite (i.e., U et V sont fortement liées dans les valeurs faibles alors qu'elles ne le sont pas dans les valeurs fortes),
- Frank : la dépendance est symétrique des deux côtés mais elle est faible sur les extrêmes et forte au centre,
- Gumbel : dépendance très forte à droite et relativement faible à gauche (l'inverse de Clayton),
- FGM : copule utile quand la dépendance entre U et V est modeste en amplitude,
- Gaussienne : copule très flexible puisqu'elle permet une dépendance positive et négative.

Hafner & Reznikova (2010) ont développé une méthode qui introduit le temps dans les formules des copules afin de créer des copules ayant une dynamique temporelle. Dans cette étude, le temps n'a pas été introduit dans les formules analytiques mais l'année de référence 2010 a été divisée en quatre saisons afin de pouvoir identifier des changements de comportement saisonniers : Jan-Fev (Décembre 2009 non disponible), Mar-Avr-Mai, Juin-Juil-Aou et Sep-Oct-Nov.

5.2.2.3 Comment choisir la famille la plus adéquate aux données

Puisque les copules séparent les distributions marginales des structures de dépendance, la copule la plus appropriée pour une application particulière est celle qui capture le mieux les caractéristiques des données (Trivedi & Zimmer, 2005). Dupuis (2007) a examiné les effets d'une mauvaise spécification et souligna les dangers d'un mauvais choix de famille de copules. La méthode de choix de la famille de copules est donc une étape qu'il est important de ne pas négliger. La méthode la plus utilisée est celle proposée par Genest & Rivest (1993) mais elle ne concerne que les copules Archimédiennes. D'autres méthodes ont été développées pour comparer tout type de copules (Fermanian, 2005, Genest et al., 2006, Genest & Rémillard, 2008, Berg, 2009). Genest et al. (2009) et Berg (2009) ont inter-comparé certaines de ces méthodes et ont conclu qu'il n'existait pas encore de méthode universelle et que certaines de ces procédures

Tableau 5.2 – Les neuf copules testées dans cette étude : définition, intervalle du paramètre θ et famille.

Copule	$C_\theta(u, v)$	$\theta \in$	Family
Indépendance	$u \cdot v$	-	-
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(0, \infty)$	Archimédienne
Frank	$-\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$	$(-\infty, \infty)/0$	Archimédienne
Gumbel	$\exp \left(- \left((-\ln u)^\theta + (-\ln v)^\theta \right)^{1/\theta} \right)$	$[1, \infty)$	Archimédienne
FGM	$uv + \theta uv(1-u)(1-v)$	$[-1, 1]$	Elliptique
AMH	$\frac{uv}{1 - \theta(1-u)(1-v)}$	$[-1, 1]$	Archimédienne
Arch12	$\left(1 + ((u^{-1} - 1)^\theta + (v^{-1} - 1)^\theta)^{1/\theta} \right)^{-1}$	$[1, \infty)$	Archimédienne
Arch14	$\left(1 + ((u^{-1/\theta} - 1)^\theta + (v^{-1/\theta} - 1)^\theta)^{1/\theta} \right)^{-1/\theta}$	$[1, \infty)$	Archimédienne
Gaussian	$\frac{\int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \exp \left(\frac{2\theta s\omega - s^2 - \omega^2}{2(1-\theta^2)} \right) ds d\omega}{2\pi\sqrt{1-\theta^2}}$	$[-1, 1]$	Elliptique

étaient meilleures pour détecter certaines formes de dépendance mais jamais pour toutes les situations.

La méthode de sélection appliquée dans cette étude est celle proposée par Huard et al. (2006) qui est basée sur une approche Bayésienne où tout type de copule peut être testé. Cette méthode ne fonctionne pas parfaitement dans certaines situations si la corrélation est très faible ou si le nombre de points testés n'est pas suffisant mais elle a l'avantage d'être une méthode très rapide comparée aux autres citées précédemment.

La méthode de Huard est basée sur le théorème de Baye pour tester si les données ont pu être générées par une copule donnée. Soient C_l la l ème copule testée ($l = 1, \dots, Q$ où Q est le nombre total de copules testées) et D les données composées de n paires mutuellement indépendantes de quantiles (u_i, v_i) avec $i = 1, \dots, n$. L'hypothèse H_l est testée : les données D ont été générées par la copule C_l . Le théorème de Baye est appliqué pour chaque famille de copule l :

$$\Pr[H_l | D, I] = \frac{\Pr[D | H_l, I] \cdot \Pr[H_l | I]}{\Pr[D | I]} \quad (5.8)$$

où I représente toute information utile supplémentaire qui peut aider au choix de la famille ; $\Pr[D | H_l, I]$ est la vraisemblance ; $\Pr[H_l | I]$ est la probabilité a priori (prior) que ce soit bien cette copule et $\Pr[D | I]$ est la constante de normalisation. La copule choisie est celle avec la plus grande probabilité $\Pr[H_l | D, I]$ (Huard et al., 2006).

5.2.2.4 Approche utilisée pour simuler des données

La clef pour générer des simulations à partir d'une fonction de copule est de comprendre qu'une copule est une distribution jointe et qu'elle obéit donc aux mêmes règles. Une distribution jointe est une probabilité qu'une première variable aléatoire soit inférieure ou égale à une première valeur et qu'une seconde variable aléatoire soit inférieure ou égale à une seconde valeur dans le même temps. Cette probabilité est également une variable aléatoire suivant une distribution uniforme \mathcal{U} sur $(0,1)$.

$$H_{XY}(x, y) = \Pr[X \leq x, Y \leq y] = p \sim \mathcal{U}(0, 1) \quad (5.9)$$

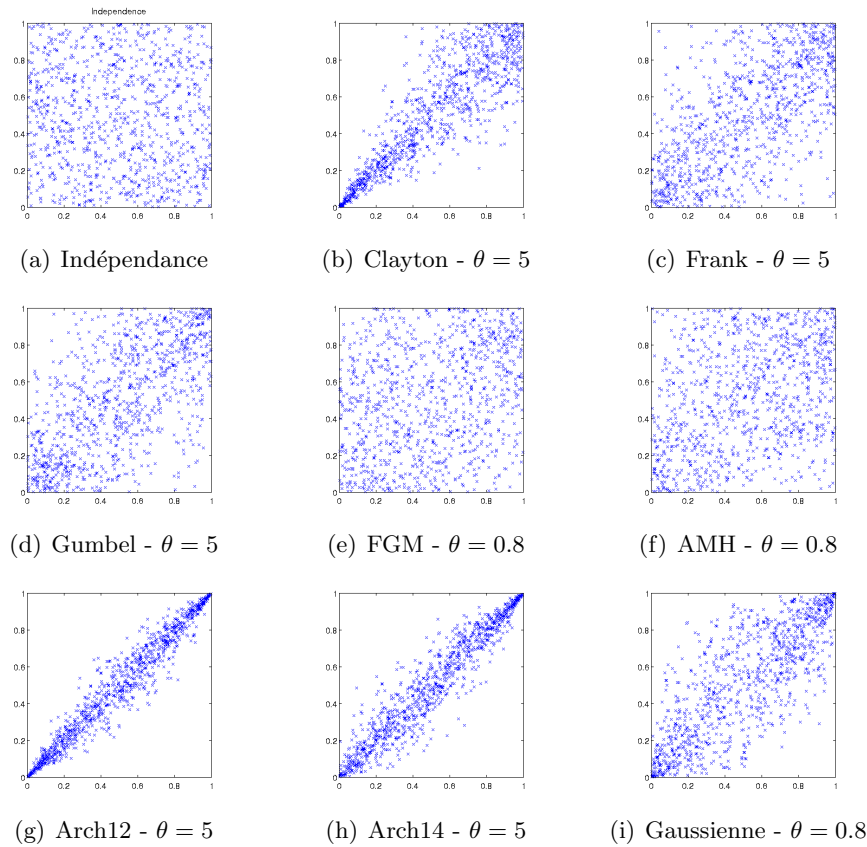


FIGURE 5.6 – Représentations des neuf copules testées dans cette étude montrant leurs caractéristiques sous forme de nuages de points (abscisses : CDF de la 1ère variable ; ordonnées : CDF de la 2ème variable).

$$C_{UV}(u, v) = \Pr[U \leq u, V \leq v] = q \sim \mathcal{U}(0, 1) \quad (5.10)$$

Une copule conditionnelle $C_{V|U}$ est la probabilité que la variable aléatoire V ait une valeur inférieure ou égale à v sachant que U est égale à la valeur u .

$$C_{V|U}(u, v) = \Pr[V \leq v \mid U = u] = t \sim \mathcal{U}(0, 1) \quad (5.11)$$

Simuler une variable uniforme t est donc nécessaire pour pouvoir générer des simulations de copules. Pour retrouver $V \mid U$, la fonction $C_{V|U}$ a besoin d'être inversée telle que $v = C_{V|U}^{-1}(t)$ ou bien que l'équation $C_{V|U}(v) = t$ soit résolue numériquement. Pour chaque valeur de t , une valeur pour v est retrouvée. L'approche suivante a donc été utilisée pour simuler des données à partir d'une fonction copule :

- calculer F_X et G_Y des deux jeux de données originaux X et Y avec l'équation (5.1)
- choisir la copule C la plus appropriée aux données en appliquant la méthode de Huard puis ajuster le paramètre θ afin d'être le plus proche des données,
- calculer la copule conditionnelle $C_{V|U}$ avec l'équation (5.5),
- générer 1000 simulations de la variable $t \sim \mathcal{U}(0, 1)$,
- calculer $v = C_{V|U}^{-1}(t)$ et $y = G_Y^{-1}(v)$, et
- calculer la moyenne et l'écart-type à partir des 1000 simulations.

Puisque la méthode des copules est une méthode statistique, il est nécessaire de générer un nombre suffisant de simulations. Dans la littérature, il n'est pas précisé le nombre minimum

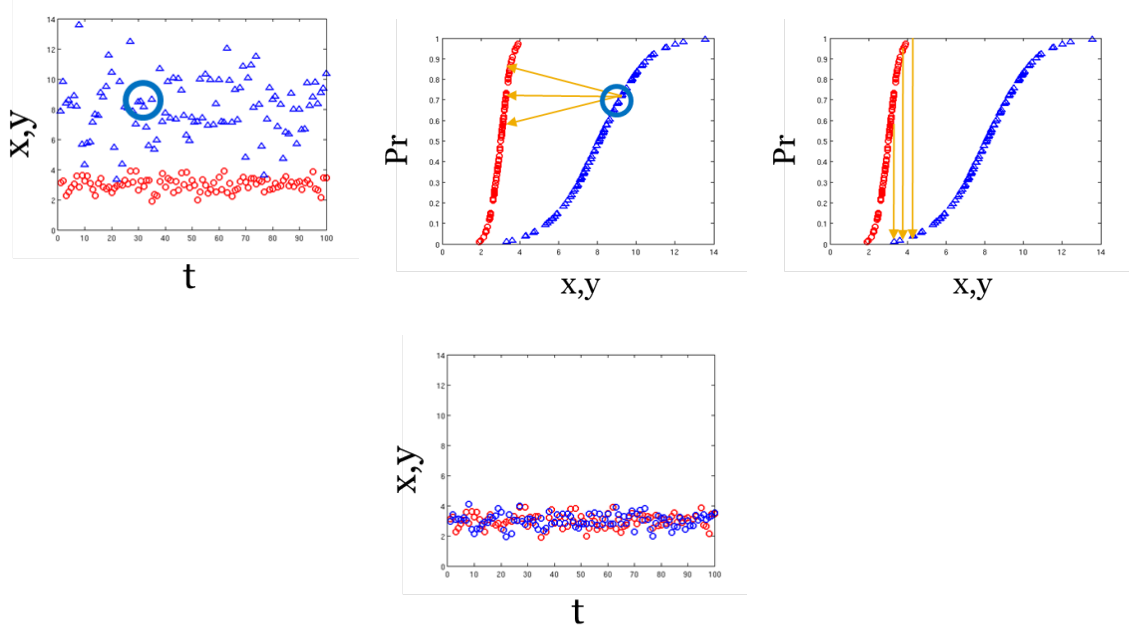


FIGURE 5.7 – Exemple d'application de la méthode des copules. La série temporelle bleue est ramenée au niveau de la série rouge en faisant correspondre leur distributions statistiques.

de simulations qui amène à des résultats cohérents donc il a été choisi arbitrairement de calculer 1000 simulations. Avec ce nombre élevé de simulations, la moyenne est statistiquement représentative et significative et peut être calculée avec un intervalle de confiance réduit.

Un autre avantage de la méthode des copules est qu'il est possible de calculer la marge d'erreur associée à la moyenne. Avec les 1000 simulations, on est en mesure de calculer l'écart-type des simulations et donc de donner une barre d'erreur aux résultats.

5.2.2.5 Exemple simple d'application

Reprenons l'exemple précédent où le but était de ramener la série temporelle bleue au niveau de la série rouge (fig. 5.7). La première étape est identique à celle du CDF matching puisqu'elle consiste à calculer les fonctions de répartition de chaque variable. On applique ensuite la méthode de Huard qui nous indique la famille de copules qui est la plus probable d'avoir pu générer de telles données. Le paramètre θ est ensuite ajusté aux données et on est en mesure de calculer la dérivée, analytiquement ou numériquement, de cette copule. Pour chaque point bleu x de la série temporelle, on obtient une probabilité u sur la distribution bleue. Grâce aux copules, pour chaque probabilité u , on génère 1000 simulations de probabilité v sur la distribution rouge qui correspondent ensuite à 1000 simulations y . Pour résumer, chaque valeur bleue x correspond à ce stade à 1000 valeurs rouges y . Une moyenne et un écart-type sont alors calculés pour ces simulations de telle façon que la série temporelle bleue est ramenée au niveau de la série rouge avec un écart-type associé.

5.2.3 Méthodologie générale

Pour les méthodes du CDF matching et des copules, les données de l'année 2010 ont été utilisées comme référence afin de calculer les distributions statistiques des humidités AMSR-E/VUA et SMOS. Les deux approches ont alors été calibrées avec 2010 puis appliquées aux années précédentes. Il est évident qu'utiliser seulement une année pour la calibration peut entraîner des dysfonctionnements, notamment si l'année 2010 ne couvre pas une gamme d'humidité suffisante par rapport aux années précédentes.

La première étape a été d'interpoler les humidités AMSR-E/VUA sur la grille de SMOS afin que les deux produits puissent être comparés sur la même grille. Afin d'avoir assez de points pour calculer des statistiques fiables, les deux méthodes ont été appliquées aux points contenus dans une région de $1^\circ \times 1^\circ$ (un peu plus de 50 points) autour des bassins versants tests présentés dans les chapitres précédents : Walnut Gulch (WG), Little Washita (LW), Little River (LR) et Reynolds Creek (RC). Seules les acquisitions du matin ont été utilisées dans cette étude : 1:30h pour AMSR-E/VUA et 6:00h pour SMOS (matin).

L'année de calibration a été divisée en quatre saisons : Décembre-Janvier-Février, Mars-Avril-Mai, Juin-Juillet-Août et Septembre-Octobre-Novembre. Cette division a été réalisée afin de mieux identifier la dynamique saisonnière qui peut être très différente à différents moments de l'année, notamment dans les zones agricoles. Avec cette division, il n'y avait pas assez de points disponibles pendant la période d'hiver pour calculer des statistiques fiables donc aucune estimation n'a été réalisée durant cette période pour aucune des deux méthodes. Dans la suite de ce chapitre, VUA représentera le produit d'humidité AMSR-E/VUA.

5.3 Séries temporelles homogènes d'humidité sur les quatre bassins versants tests aux USA

L'année 2010 est utilisée pour calculer les distributions statistiques des produits d'humidité de SMOS et de VUA pour les deux méthodes testées : les CDF matching et les copules. Des valeurs d'humidité sont estimées en utilisant la distribution conditionnelle, conditionnées par les valeurs de VUA. Les simulations des copules permettent d'avoir un ensemble de valeurs que l'on représente ici par leur moyenne et leur écart-type. Le CDF matching ne fournit qu'une seule valeur d'humidité sans incertitude associée. La série temporelle ainsi générée sera donc homogène dans le temps sous l'hypothèse que les humidités VUA et que les températures de brillances AMSR-E soient temporellement cohérentes (pas de biais interne). Ces séries temporelles sont ensuite comparées aux mesures d'humidité du sol sur les quatre bassins versants test.

5.3.1 Application sur l'année 2010 des deux méthodes

Les données de 2010, qui est l'année commune entre SMOS et VUA, ont été utilisées pour calculer les distributions statistiques des deux variables. Les méthodes de CDF matching et des copules ont ensuite été appliquées et ont produit des estimations d'humidité différentes. Sur la figure 5.8, les données d'origine SMOS et VUA sont représentées par le nuage de points bleus avec les simulations par CDF matching en vert et par copules en violet. L'écart-type calculé à partir des simulations des copules peut être interprété comme l'incertitude de ces simulations ou comme un indicateur de qualité. Cette incertitude ne peut pas être fournie par le CDF matching qui ne simule qu'une seule valeur et non un ensemble.

Sur le site de Walnut Gulch (WG), pendant la période Mars-Avril-Mai (MAM), il n'y a pas de différence notable entre les deux simulations, elles restent très proches des données d'origine. Cependant pendant les saisons Juin-Juillet-Août (JJA) et Septembre-Octobre-Novembre (SON),

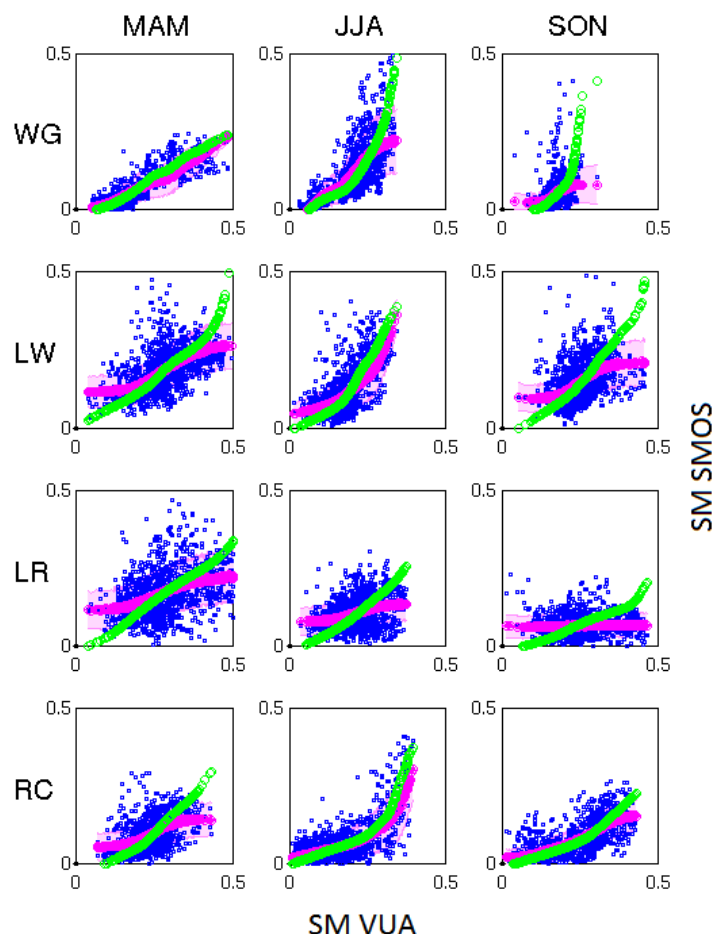


FIGURE 5.8 – Différences de simulations des humidités (SM en m^3/m^3) entre le CDF matching et la méthode des copules. Les données d'origine sont en points bleus, les données simulées par CDF matching en vert et celles simulées par copules en rouge. L'incertitude de la méthode des copules est indiquée en ombre rouge. Chaque ligne correspond à un site test et chaque colonne à une saison de l'année.

des différences apparaissent pour les hautes valeurs d'humidité où le CDF matching produit des valeurs plus hautes que celles simulées par copules. Le même phénomène est visible durant toutes les saisons sur les trois autres sites, Little Washita (LW), Little River (LR) et Reynolds Creek (RC). Des différences sont également visibles dans les faibles valeurs d'humidité sur LW, LR et RC (MAM) où les copules génèrent de plus grandes valeurs d'humidité que le CDF matching.

Les écart-types ont également été calculés à partir des simulations des copules (fig. 5.8). Cette incertitude est directement reliée à la largeur du nuage de points de la copule qui est en réalité contrôlée par le paramètre θ . Une valeur élevée de l'incertitude correspond à un nuage de points plus large, signifiant que les deux variables sont faiblement liées, alors qu'une faible incertitude correspond à un lien fort entre les variables.

Les différences de simulations peuvent également être observées dans les séries temporelles de 2010 (fig. 5.9 et tab. 5.3). Comparées aux données VUA d'origine, les humidités simulées sont proches du niveau de SMOS et comparables aux mesures terrain. Le biais qui existait entre SMOS et VUA a été correctement corrigé par les deux méthodes.

Sur Walnut Gulch, les simulations par CDF matching et par copules ne sont pas très différentes à part en été (fig. 5.9(a)) où les simulations par CDF matching sont plus hautes, notamment après un évènement pluvieux (exemple du mois d'Août). En considérant l'année entière, les deux méthodes améliorent les statistiques d'origine du produit VUA. Le coefficient de corrélation ne change pas significativement ($R=0.79$ pour VUA et $R=0.79/0.82$ pour les copules/CDF matching) mais l'erreur totale RMSE est grandement améliorée passant de $0.139 \text{ m}^3/\text{m}^3$ (VUA original) à $0.054 \text{ m}^3/\text{m}^3$ avec le CDF matching et $0.043 \text{ m}^3/\text{m}^3$ avec les copules, ce qui représente une amélioration d'un facteur 3.

Sur le bassin versant de Little Washita, les simulations répondent très bien aux différents évènements pluvieux successifs tout au long de l'année et ont une très bonne dynamique dans la phase de décroissance après une pluie. Les simulations des deux premiers mois (Mars-Avril, fig. 5.9(b)) sont un peu plus bruitées et cela impacte les statistiques avec une corrélation de $0.55/0.57$ et une erreur de $0.057/0.075 \text{ m}^3/\text{m}^3$ pour les copules/CDF matching. Les deux autres saisons donnent de très bons résultats statistiques (tab. 5.3). Quand on considère l'année entière, la corrélation a été grandement améliorée : $R=0.59$ pour les données VUA d'origine et $R=0.71/0.71$ pour les simulations par copules/CDF matching. L'erreur RMSE a été quant à elle réduite d'un facteur 3 avec $\text{RMSE}=0.148 \text{ m}^3/\text{m}^3$ pour VUA et $\text{RMSE}=0.043/0.059 \text{ m}^3/\text{m}^3$ pour les copules/CDF matching.

Little River est le site avec la plus grande fréquence de pluies même si elles ne sont que de faible amplitude. Ces évènements pluvieux successifs ne sont pas bien retranscrits par les simulations, plus spécialement pendant l'automne où les deux simulations ne montrent que de faibles variations d'amplitude (fig. 5.9(c)), ce qui a entraîné de très mauvais résultats statistiques ($R=0.17/0.16$ pour les copules/CDF matching). Malheureusement, même si les pluies ont été capturées par SMOS et VUA, aucune pluie n'a été simultanément capturée par ces deux produits d'origine. De ce fait, les relations établies entre SMOS et VUA ne peuvent pas retranscrire correctement ces évènements. Il y a sur ce site un problème de représentativité. Les simulations ne peuvent donc être représentatives que des périodes sèches. On peut noter également que les statistiques de VUA ne sont déjà pas très bonnes pendant la période sèche de Septembre à Novembre avec $R=0.37$ et $\text{RMSE}=0.174 \text{ m}^3/\text{m}^3$. Durant la saison du printemps, l'humidité SMOS est sur-estimée, ce qui a pour conséquence de sur-estimer l'humidité dans les simulations des deux méthodes comparées aux mesures terrain.

Reynolds Creek est situé dans une région montagneuse et subit donc de nombreuses chutes de neige et de gels des sols. De ce fait, les humidités retrouvées par satellites ne sont pas comparables avec les mesures terrain avant fin Mai. Après cette période, les simulations retranscrivent précisément l'évolution de l'humidité et améliorent les statistiques d'origine et plus spécialement l'erreur, $\text{RMSE}=0.099 \text{ m}^3/\text{m}^3$ pour VUA et $\text{RMSE}=0.059/0.067 \text{ m}^3/\text{m}^3$ pour les simulations par copules/CDF matching. Il faut tout de même noter que la variabilité spatiale des données terrain est grande sur ce site (fig. 5.9(d)).

5.3.2 Application aux données 2003-2010

L'humidité de 2003 à 2010 est simulée à partir des valeurs retrouvées par VUA en utilisant la méthode des copules et le CDF matching, toutes deux calibrées avec les données de l'année 2010. La figure 5.10 et le tableau 5.4 présentent ces séries temporelles homogènes et les statistiques associées (R et RMSE) entre les données d'origine VUA, les simulation par copules, les simulations par CDF matching et les mesures terrain.

Walnut Gulch est le site le plus sec de cette étude et n'a eu que très peu d'évènements pluvieux entre 2003 et 2010. Les quelques pluies sont cependant bien décrites par les simulations même si elles sont parfois largement sur-estimées par le CDF matching. Des artefacts sont visibles aux limites des saisons au début des années 2006 et 2008. Le coefficient de corrélation

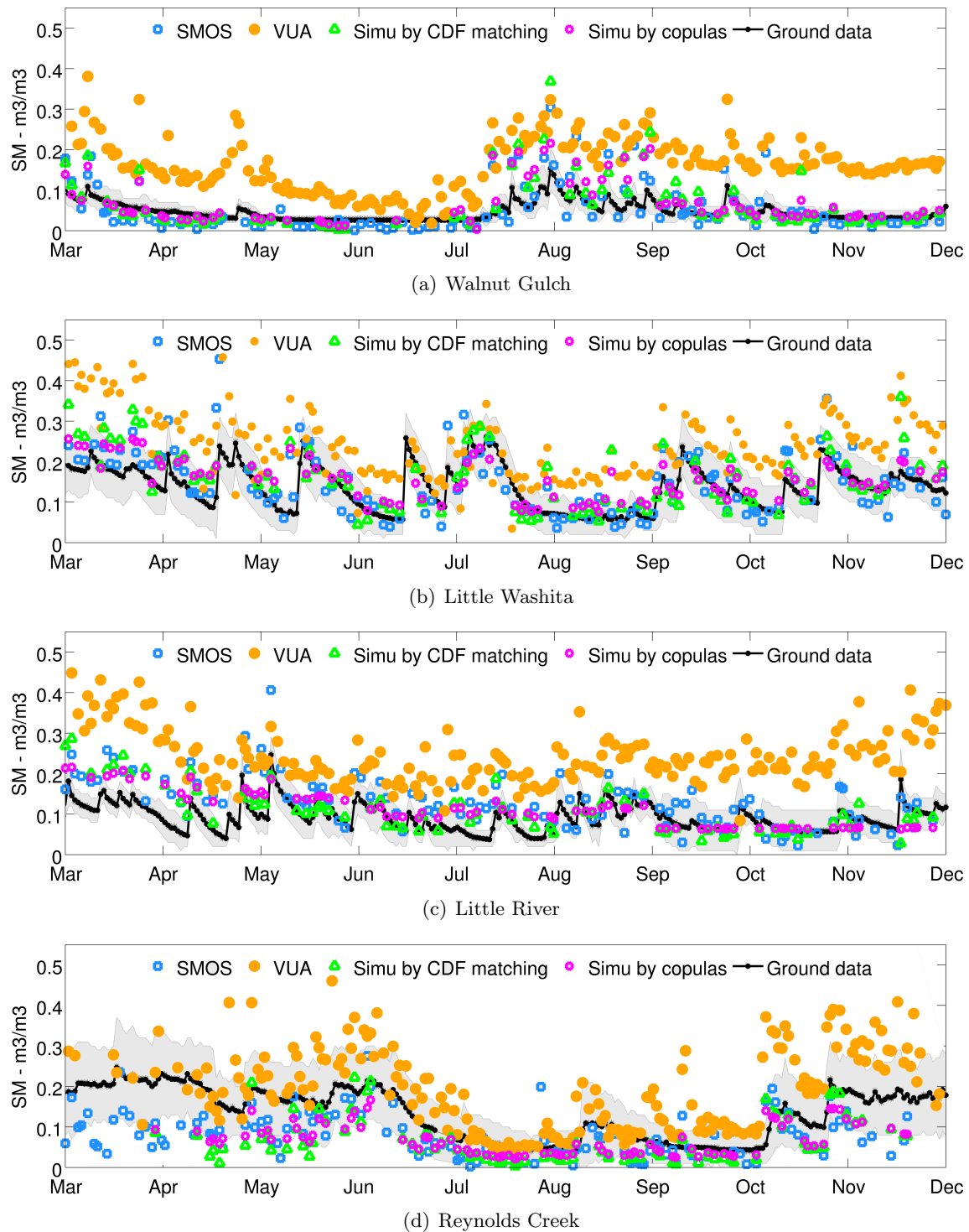


FIGURE 5.9 – Simulations pour l'année 2010 : SMOS, VUA, humidités simulées par CDF matching et copules et les mesures terrain sur les quatre bassins versants. Puisque les données terrain sont des moyennes de plusieurs stations, leur écart-type est représenté en ombre noire montrant ainsi la variabilité spatiale à l'intérieur de chaque bassin versant.

Tableau 5.3 – Résultats statistiques des simulations des copules et du CDF matching. Ces simulations sont comparées aux mesures terrain d'humidité en 2010 divisées en 4 saisons : Mars-Avril-Mai (MAM), Juin-Juillet-Août (JJA), Septembre-Octobre-Novembre (SON), mais il n'y avait pas assez de données disponibles pour la saison d'hiver. Les résultats significativement meilleurs sont indiqués en gras et les RMSE sont en m^3/m^3 .

		SMOS		VUA		Copules			CDF matching		# dts
		R	RMSE	R	RMSE	Fam(θ)	R	RMSE	R	RMSE	
WG	MAM	0.80	0.032	0.82	0.125	Gumbel (2.18)	0.89	0.020	0.87	0.031	43
	JJA	0.86	0.053	0.86	0.126	Clayton (2.63)	0.76	0.076	0.81	0.090	45
	SON	0.64	0.029	0.79	0.133	Frank (3.13)	0.64	0.012	0.53	0.029	42
	total	0.84	0.040	0.79	0.139	-	0.79	0.043	0.82	0.054	130
LW	MAM	0.70	0.068	0.48	0.166	Frank (4.40)	0.55	0.057	0.57	0.075	44
	JJA	0.85	0.037	0.58	0.085	Gumbel (1.66)	0.77	0.042	0.76	0.050	44
	SON	0.80	0.041	0.80	0.122	Frank (3.61)	0.75	0.023	0.72	0.048	46
	total	0.78	0.049	0.59	0.148	-	0.71	0.043	0.71	0.059	134
LR	MAM	0.77	0.080	0.54	0.175	Frank (2.82)	0.59	0.063	0.58	0.067	39
	JJA	0.57	0.053	0.67	0.131	Frank (2.00)	0.65	0.034	0.66	0.033	40
	SON	0.59	0.032	0.37	0.174	FGM (0.31)	0.17	0.033	0.16	0.037	39
	total	0.74	0.060	0.65	0.178	-	0.51	0.045	0.59	0.048	118
RC	MAM	0.14	0.097	0.11	0.096	Frank (3.10)	0.26	0.089	0.27	0.105	47
	JJA	0.63	0.055	0.81	0.070	Gumbel (1.81)	0.84	0.047	0.83	0.052	42
	SON	0.14	0.070	0.52	0.144	Frank (6.30)	0.34	0.056	0.29	0.066	39
	total	0.55	0.081	0.73	0.099	-	0.80	0.059	0.70	0.067	128

est globalement meilleur avec le CDF matching alors que l'erreur est divisée par au moins 2 avec les simulations par copules.

La sur-estimation des événements pluvieux par le CDF matching peut aussi être trouvée sur le site de Little Washita. L'évolution temporelle de l'humidité est cependant bien redessinée par les deux méthodes. Sur ce site, le CDF matching sur-estime les hautes humidités alors qu'elle sous-estime les faibles humidités, signifiant que le CDF matching reproduit l'humidité avec une amplitude bien plus grande que les mesures terrain et que les simulations par copules. Cela se reflète dans les valeurs des erreurs du CDF matching, $RMSE=0.079 \text{ m}^3/\text{m}^3$ alors que les simulations par copules obtiennent une $RMSE$ de $0.066 \text{ m}^3/\text{m}^3$ ($RMSE$ des données VUA d'origine : $0.160 \text{ m}^3/\text{m}^3$).

Little River est le site avec le plus grand nombre d'événements pluvieux et cette haute fréquence de pluie n'est pas capturée proprement pendant la saison d'automne 2010. Ce manque se retranscrit donc sur toute la série temporelle de 2003 à 2010 où les simulations sont complètement plates et sans aucune dynamique durant les saisons d'automne. De plus, puisque SMOS sur-estime l'humidité pendant le printemps 2010, les deux méthodes statistiques ont ce même comportement. Bien que la tendance générale des simulations soit correcte, la dynamique n'est pas bien représentée et cela se traduit par de mauvaises performances statistiques avec des coefficients de corrélation parfois négatifs (2004, 2007, tab. 5.4).

Reynolds Creek est un site compliqué à plusieurs niveaux : le climat, avec beaucoup de chutes de neige et de gel, et la topographie car ce site est situé dans les montagnes. Cependant, les résultats statistiques sont améliorés par les simulations par copules avec une erreur totale quasiment divisée par 2 : $RMSE=0.099 \text{ m}^3/\text{m}^3$ pour les données d'origine VUA et $0.056/0.062 \text{ m}^3/\text{m}^3$ pour les copules/CDF matching.

5.4 Développements pour étendre ces méthodes à l'échelle globale

Afin d'étendre cette méthodologie à l'échelle globale, une étude sur le possible lien entre l'occupation des sols et le choix de la famille de copules a été menée. Les résultats qui suivent sont des résultats préliminaires et nécessiteraient d'être approfondis.

Dans cette optique de lien possible entre l'occupation du sol et le choix de la famille de copules entre SMOS et VUA, certaines zones homogènes ont été sélectionnées afin de s'affranchir de l'hétérogénéité qui pourrait introduire du bruit. Des zones de prairies et de cultures ont ainsi été sélectionnées (min. 80%) au centre et au Nord des États-Unis (figs. 5.11 et 5.12). Les humidités SMOS et VUA ont été collectées sur ces points puis la méthode de Huard a été appliquée pour choisir la famille de copules la plus adaptée. Cette méthode fournit également les poids associés à chaque famille testée et celle qui obtient le plus grand poids est celle qui est choisie. Cependant, le poids de la famille sélectionnée pour représenter les données n'est pas tout le temps de 1 (valeur maximale) et il peut subsister des doutes. Sur la figure 5.12, on identifie clairement les zones de transition entre les choix avec ce poids qui diminue aux limites des zones couvrant un même choix.

Concernant les zones de prairies, la famille de copules Frank a été la plus choisie. Elle représente une dépendance forte pour les valeurs modérées et faible pour les valeurs faibles et fortes (fig. 5.6(c), p.95). La deuxième copule la plus représentée est Gumbel qui traduit une forte dépendance dans les valeurs élevées et une faible dépendance pour les petites valeurs d'humidité. Même si pour le moment, aucune explication à ce découpage n'a été clairement identifiée, les familles de copules n'ont pas été choisies aléatoirement et forment des regroupements et des structures spatiales qui doivent être reliées à des comportements des humidités (SMOS et VUA) bien distincts. Une autre piste pourrait être celle de la texture du sol.

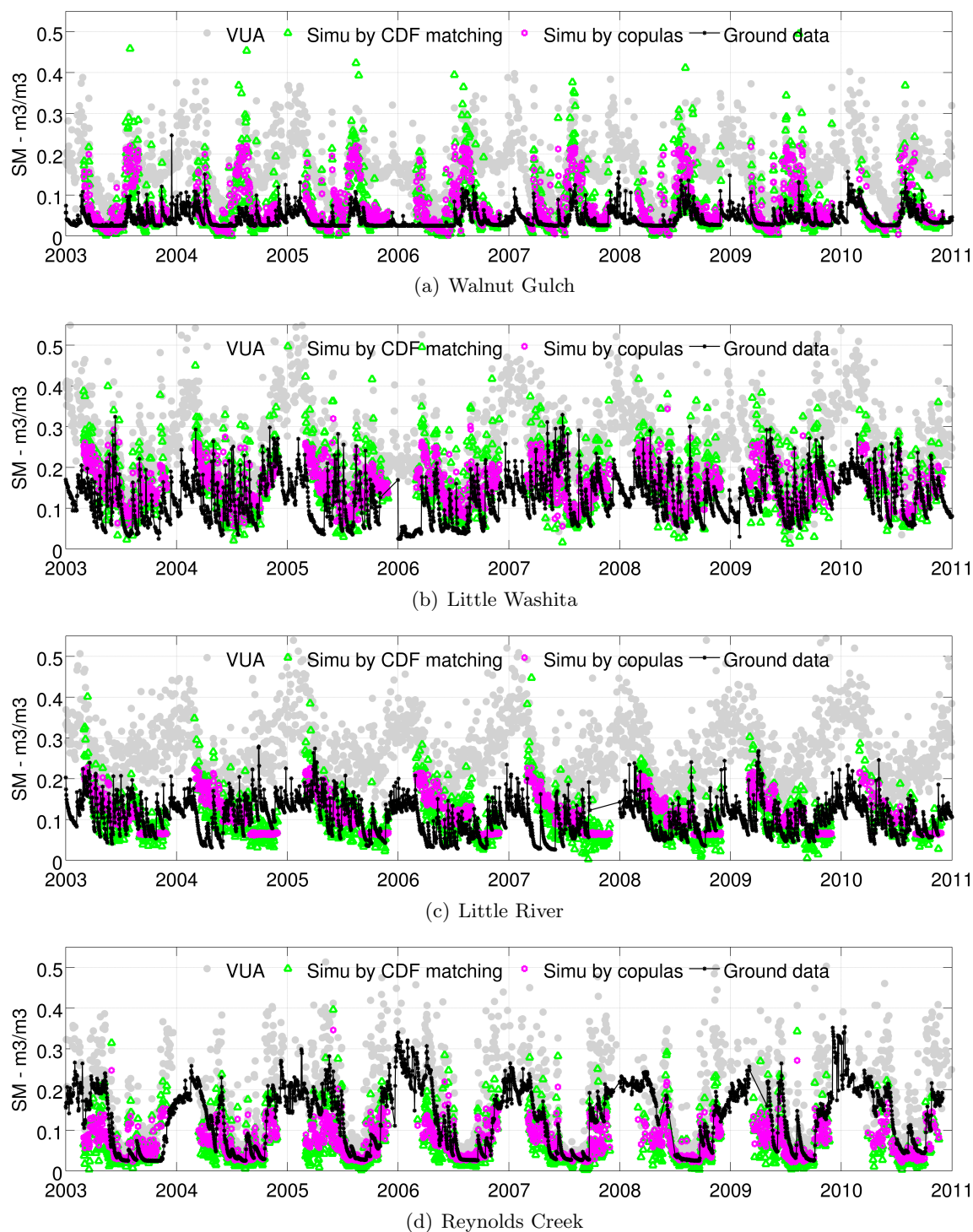


FIGURE 5.10 – Séries temporelles homogènes des humidités des sols (SM en m^3/m^3) simulées par CDF matching (vert) et par copules (violet) de 2003 à 2010. Les mesures terrain sur les 4 bassins versants sont indiquées en noir alors que les données VUA d'origine sont en gris. L'hiver ne faisant pas partie de l'étude, aucune simulation n'a été réalisée sur cette saison.

Tableau 5.4 – Résultats statistiques de la comparaison entre les séries temporelles simulées de l'humidité de 2003 à 2010. L'historique d'origine est représenté par VUA. Les résultats significativement meilleurs sont indiqués en gras et les RMSE sont en m^3/m^3 .

(a) Walnut Gulch										
		2003	2004	2005	2006	2007	2008	2009	2010	Total
VUA	R	0.70	0.76	0.82	0.66	0.81	0.68	0.65	0.79	0.73
	RMSE	0.129	0.141	0.146	0.133	0.147	0.138	0.129	0.139	0.138
Cop.	R	0.62	0.55	0.82	0.64	0.81	0.75	0.76	0.79	0.69
	RMSE	0.059	0.059	0.059	0.060	0.054	0.053	0.060	0.043	0.057
CDF	R	0.73	0.62	0.88	0.72	0.89	0.75	0.79	0.82	0.74
	RMSE	0.070	0.074	0.071	0.073	0.067	0.067	0.077	0.054	0.071
(b) Little Washita										
		2003	2004	2005	2006	2007	2008	2009	2010	Total
VUA	R	0.56	0.71	0.48	0.67	0.32	0.42	0.52	0.58	0.55
	RMSE	0.163	0.149	0.187	0.149	0.173	0.158	0.149	0.149	0.160
Cop.	R	0.56	0.47	0.19	0.62	0.41	0.64	0.58	0.71	0.47
	RMSE	0.071	0.064	0.088	0.077	0.060	0.056	0.051	0.044	0.066
CDF	R	0.59	0.60	0.34	0.63	0.49	0.61	0.53	0.71	0.51
	RMSE	0.083	0.070	0.101	0.092	0.069	0.076	0.069	0.059	0.079
(c) Little River										
		2003	2004	2005	2006	2007	2008	2009	2010	Total
VUA	R	0.51	0.60	0.46	0.75	0.64	0.70	0.49	0.65	0.58
	RMSE	0.171	0.148	0.181	0.185	0.180	0.166	0.187	0.178	0.174
Cop.	R	0.54	-0.48	0.73	0.01	-0.14	0.20	0.43	0.51	0.19
	RMSE	0.042	0.079	0.036	0.069	0.081	0.054	0.047	0.045	0.059
CDF	R	0.68	-0.16	0.72	0.28	0.18	0.50	0.55	0.59	0.37
	RMSE	0.044	0.080	0.042	0.070	0.085	0.050	0.048	0.048	0.061
(d) Reynolds Creek										
		2003	2004	2005	2006	2007	2008	2009	2010	Total
VUA	R	0.78	0.76	0.74	0.80	0.84	0.69	0.78	0.73	0.77
	RMSE	0.093	0.085	0.110	0.099	0.102	0.106	0.099	0.099	0.099
Cop.	R	0.53	0.78	0.70	0.68	0.72	0.75	0.72	0.80	0.69
	RMSE	0.065	0.045	0.065	0.060	0.051	0.047	0.052	0.059	0.056
CDF	R	0.42	0.69	0.65	0.63	0.70	0.65	0.71	0.70	0.63
	RMSE	0.073	0.051	0.070	0.063	0.055	0.056	0.056	0.067	0.062

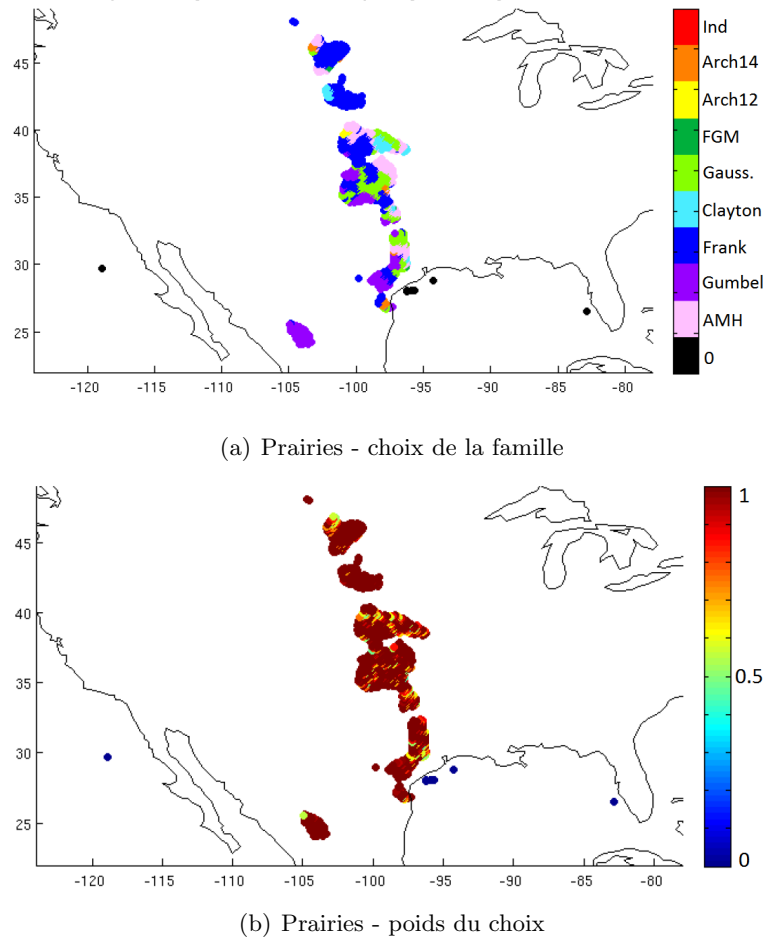


FIGURE 5.11 – Choix de la famille de copules la plus adaptée (a) pour faire un lien entre les humidités SMOS et VUA sur les zones de prairies pour les mois d’Avril, Mai et Juin. La méthode de Huard fournit également un coefficient de poids (b) permettant de juger ce choix : 1 pour une certitude, 0 quand le choix est très incertain.

Pour les zones de cultures, Gumbel a été la copule la plus choisie suivie de la copule de Clayton. Ces deux copules représentent des structures de dépendance opposées puisque Gumbel modélise une forte dépendance des hautes valeurs (fig. 5.6(d), p.95) alors que Clayton modélise une forte dépendance des faibles valeurs d’humidité (fig. 5.6(b)). De la même manière que pour les prairies, on observe clairement des formes spatiales où le choix de la famille est le même. Une cause pourrait être les différents stades de croissance dans lesquels se trouvent les cultures observées.

5.5 Conclusions

Le principal objectif de ce chapitre était de proposer une nouvelle méthode pour générer une longue série temporelle de l’humidité des sols de 2003 à 2010 à partir de deux produits d’humidité.

Dans ce but, deux outils statistiques ont été utilisés et testés sur quatre bassins versants aux États-Unis : le CDF matching et la méthode des copules. En utilisant le CDF matching, l’hypothèse posée est que les deux jeux de données sont ordonnés de la même manière (même chronologie) alors que les copules ne nécessitent pas cette hypothèse. Les deux jeux de données

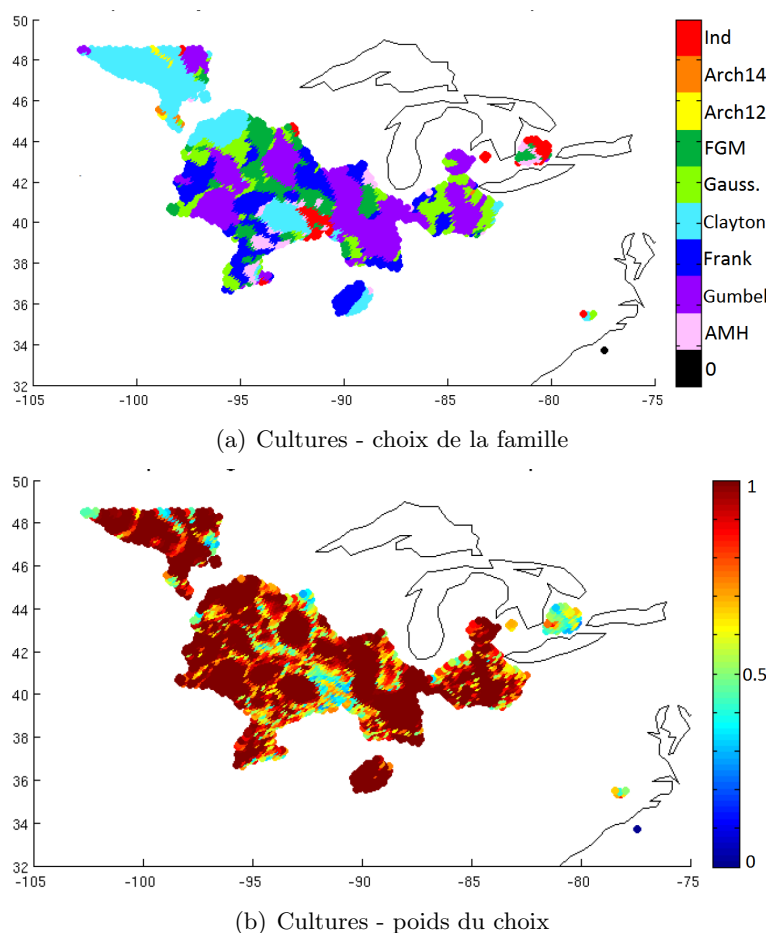


FIGURE 5.12 – Choix de la famille de copules la plus adaptée (a) pour faire un lien entre les humidités SMOS et VUA sur les zones de cultures pour les mois d'Octobre, Novembre et Décembre. La méthode de Huard fournit également un coefficient de poids (b) permettant de juger ce choix : 1 pour une certitude, 0 que le choix n'est pas sûr du tout.

d'origine, SMOS et AMSR-E/VUA, étaient disponibles conjointement durant l'année 2010, donc cette année 2010 a été utilisée comme référence par les deux méthodes statistiques, ce qui a permis de calculer les distributions de chaque variable afin d'appliquer ensuite ces méthodes aux années précédentes. La nouveauté de l'approche réside dans son application : établir une relation statistique entre les humidités de VUA et de SMOS afin d'estimer des *équivalents* SMOS pendant la période d'existence de VUA.

La première analyse des simulations en 2010 a montré que les humidités simulées étaient similaires aux humidités de SMOS et reproduisaient la dynamique de SMOS sauf sur Little River où de nombreux événements pluvieux se sont produits. Cette haute fréquence de pluies a été interprétée statistiquement comme du bruit et de ce fait, les simulations ne décrivaient pas correctement l'évolution temporelle de l'humidité sur ce site. Reynolds Creek était également un site compliqué à cause de sa topographie et de ses conditions climatiques saisonnières comme la neige et le gel des sols. Les humidités retrouvées à partir d'observations de satellites n'étaient pas capables de reproduire précisément la dynamique de l'humidité comme dessinée par les mesures terrain et par conséquent, les simulations ont échoué. Cependant, l'erreur totale des simulations par copules a été réduite d'un facteur 2. Les humidités de Walnut Gulch et Little Washita étaient très bien reproduits par les simulations et l'utilisation des copules a permis

d'améliorer l'erreur totale d'un facteur 3 tout en obtenant un niveau de corrélation équivalent aux simulations par CDF matching.

Des séries temporelles de 2003 à 2010 ont été simulées et comparées aux mesures terrain sur les quatre bassins versants. Sur Little River, les humidités simulées en 2010 avait une dynamique faible et cela est resté le cas pour la série temporelle entière, ce qui s'est traduit par des résultats statistiques assez mauvais en terme de corrélation, même si l'erreur totale a été divisée par 3. Pour les trois autres sites, le coefficient de corrélation a été légèrement dégradé comparé au jeu de données originales VUA, mais l'erreur totale a été améliorée d'un facteur 2 à 3. En général, le CDF matching a fourni de meilleures corrélations alors que les copules de plus basses erreurs tout en restant très proches des valeurs de corrélation du CDF matching.

En conclusion plus générale, le CDF matching donne de bons résultats mais ne prend pas en compte la structure de la dépendance entre les deux jeux de données d'origine (une dépendance chronologique est supposée) alors que les copules permettent de modéliser cette structure. A travers le choix de la famille et du paramètre θ (qui contrôle la largeur du nuage de points), il est possible de modéliser toute sorte de structure, de la dépendance parfaite (CDF matching) à l'indépendance complète en passant par une dépendance uniquement des fortes ou des faibles valeurs. Pour ces raisons, les copules fournissent de meilleurs résultats que le CDF matching aux valeurs extrêmes : très basses ou très hautes, comme après une pluie par exemple.

Les copules peuvent également estimer l'incertitude de ses simulations qui peut être interprétée comme un indicateur de qualité des simulations. Grâce à la méthode de Huard, il est très rapide de choisir la famille de copules qui se rapproche le plus des données ainsi que son paramètre associé puisque cette méthode est basée sur une approche Bayésienne. Les 1000 simulations sont en revanche longues, d'autant plus lorsqu'il est nécessaire de résoudre numériquement l'équation avec la copule conditionnelle. Néanmoins, ces simulations représentent un avantage certain puisqu'il est alors possible de calculer leur moyenne et leur écart-type qui sert d'indicateur d'incertitude. Ces incertitudes sont appréciées dans des modèles d'assimilation de données.

Les limitations sont les mêmes pour les deux méthodes statistiques présentées dans ce chapitre et pour toutes les méthodes en général qui utilisent une année comme référence pour leur calibration : seule la dynamique de l'année de référence peut être reproduite. Cela signifie que si un événement extrême se produit dans les années précédentes et qu'il n'est pas représenté dans l'année de référence (sécheresse exceptionnelle ou inondation) alors cet événement sera peut être mal représenté par ces méthodes. Dans cette étude, le cas de Little River en est l'exemple durant la saison d'automne.

Dans le but d'améliorer cette méthodologie, appliquer une fenêtre temporelle glissante de trois mois au lieu de découper l'année en saisons permettrait de s'affranchir des problèmes de transition entre les saisons. Une autre solution serait d'introduire le temps dans les formules de copules mais le niveau de complexité dans la manipulation de ces fonctions deviendrait trop important, plus spécialement pour les copules conditionnelles.

Le but final serait de pouvoir construire une série temporelle de l'humidité des sols à l'échelle globale. Pour cela il faut persévérer dans l'étude du choix de copules en fonction du type de sol, de la couverture du sol ou de tout autre paramètre physique, ce qui permettrait de s'affranchir de la phase du choix de la famille de copules à partir des données d'origine.

Cette étude a été menée en collaboration avec le Professeur Eric Wood de l'université de Princeton aux Etats-Unis et a abouti à un article qui a été soumis en Janvier 2012 au journal *IEEE Transactions on Geoscience and Remote Sensing* se nommant *An approach to constructing a homogeneous time series of soil moisture using SMOS*. Les auteurs de cet article sont Delphine J. Leroux, Yann H. Kerr (CESBIO, Toulouse), Eric F. Wood, Alok Sahoo (Université de Princeton, États-Unis), Rajat Bindlish et Tom J. Jackson (USDA, Beltsville, États-Unis). J'ai également pu présenter cette étude sous forme d'oral à la conférence des *IGARSS (Internation*

tional Geoscience And Remote Sensing Symposium, Vancouver, Canada) en 2011 et sous forme de posters à *RAQRS* (*Recent Advances in Quantitative Remote Sensing*, Valence, Espagne) en 2010 et *SPAT* (*Spatial Statistics*, Enschede, Pays-Bas) en 2011.

Chapitre 6

Création d'un algorithme simplifié SMOS

De nombreuses observations provenant de différents satellites ont déjà permis d'estimer l'humidité à l'échelle globale. Parmi les modélisations existantes, le modèle LPRM (*Land Parameter Retrieval Model*, Owe et al. (2001)) a été appliqué aux acquisitions des satellites SMMR, SSM/I, AMSR-E et a ainsi pu fournir un historique de l'humidité des sols des trente dernières années. Depuis son lancement en Novembre 2009, le satellite SMOS fait maintenant partie de cette série de satellites qui permettent de retrouver l'humidité et il est important d'inclure ces nouvelles acquisitions dans cet historique.

Dans la première partie de chapitre, une description détaillée du modèle LPRM est réalisée afin de bien comprendre les éléments essentiels et les enjeux de l'adaptation de ce modèle aux acquisitions multi-angulaires en bande L de SMOS. Les différentes phases de cette adaptation et les options envisagées sont ensuite décrites. Enfin toutes les approches sont testées sur deux parties d'orbite aux États-Unis et en Australie pour étudier la distribution spatiale des humidités retrouvées et sur l'année 2010 sur les bassins versants de Little Washita et de Walnut Gulch pour couvrir l'aspect temporel.

Le modèle LPRM adapté avec les optimisations de l'humidité et de l'albédo ainsi que l'introduction de la dépendance de la rugosité à l'humidité du sol est le modèle qui fournit les valeurs d'humidité les plus proches des mesures terrain et des humidités SMOS.



6.1 Introduction

6.1.1 Contexte

SMOS (*Soil Moisture and Ocean Salinity*) est la première mission satellite entièrement conçue pour retrouver l'humidité des sols sur les terres émergées (Kerr et al., 2001, 2010). Plusieurs aspects novateurs ont été introduits dans cette mission, au niveau de la technique d'acquisition aussi bien que dans la méthode d'inversion de l'humidité. On parle d'inversion car le signal acquis est utilisé pour retrouver un paramètre qui a contribué à ce signal.

Un des aspects novateurs de la mission SMOS porte sur la méthode pour retrouver l'humidité des sols où la diversité de la surface observée est prise en compte. En effet, le radiomètre a une résolution de 40 km mais le signal observé par l'instrument provient en réalité d'une zone bien plus large. Pour tenir compte de cette zone d'influence, une fonction de pondération (qui dépend du diagramme d'antenne et de l'angle d'incidence) est appliquée. Pour prendre en compte la diversité de la surface, une zone de travail ou *working area* (présentée au chapitre 2) est définie autour de chaque point. Cette zone de travail est ensuite fragmentée en plusieurs classes d'occupation des sols (forêt, végétation, ville, eau douce, sol aride, glace, eau salée, zone humide) (ATBD, 2010) qui sont définies par ECOCLIMAP (Masson et al., 2003). Afin de ne retrouver l'humidité que sur les surfaces où cela a du sens (forêt, végétation), la contribution au signal des autres fractions est estimée à l'aide de données extérieures comme ECMWF (*European Center for Medium range Weather Forecast*) par exemple. Il est alors possible de retrouver les informations d'humidité et de végétation uniquement sur les fractions d'intérêt.

Cette prise en compte de la diversité de la zone observée a été réalisée pour la première fois avec la mission SMOS. Auparavant, les humidités telles que celles retrouvées à partir des observations du radiomètre AMSR-E (*Advanced Microwave Scanning Radiometer Earth*) ou du radar ASCAT (*Advanced Scatterometer*) sont retrouvées à partir de la totalité du signal acquis, comme si la surface observée était uniforme (fig. 2.8, p. 21). De plus, aucun satellite n'avait encore fait d'acquisition multi-angulaire dans cette bande de fréquence.

Dans un contexte d'étude climatique, des historiques homogènes sont requis et l'utilisation d'un algorithme unique serait une première étape.

6.1.2 Objectifs

Il existe un historique de l'humidité des trente dernières années, recréé par l'université d'Amsterdam, Pays-Bas. L'algorithme LPRM (*Land Parameter Retrieval Model*, Owe et al. (2001)) a été utilisé pour cet historique et a été adapté pour retrouver l'humidité de surface et l'épaisseur optique de la végétation à partir d'acquisitions micro-ondes. Cette série temporelle est composée de plusieurs capteurs : SMMR (*Scanning Multi-channel Microwave Radiometer*), la série des SMM/I (*Special Sensor Microwave/Imager*), TRMM-TMI (*Tropical Rainfall Measuring Mission - Microwave Imager*) et AMSR-E. Cet algorithme s'appuie sur la bande C de SMMR et d'AMSR-E, la bande X de SMMR, de TRMM et d'AMSR-E et la bande Ka de SSM/I. La figure 2.3 montre bien que ces fréquences ne sont pas les plus adaptées pour retrouver l'humidité mais elles étaient les seules disponibles jusqu'au lancement de SMOS en 2009.

L'objectif de ce chapitre est d'adapter l'algorithme LPRM aux acquisitions multi-angulaires de SMOS en bande L afin d'inclure les acquisitions SMOS dans cette longue série temporelle de l'humidité. Pour cela, il faut réutiliser le même algorithme. La première étape est donc l'adaptation de cet algorithme à travers la calibration de certaines constantes ainsi que la modification de l'optimisation afin de prendre en compte tous les angles. L'étape suivante est de valider spatialement et temporellement ce nouveau produit d'humidité. Il sera ainsi possible de comparer les variations spatiales et temporelles avec les produits d'humidité de SMOS et d'AMSR-E. Pour

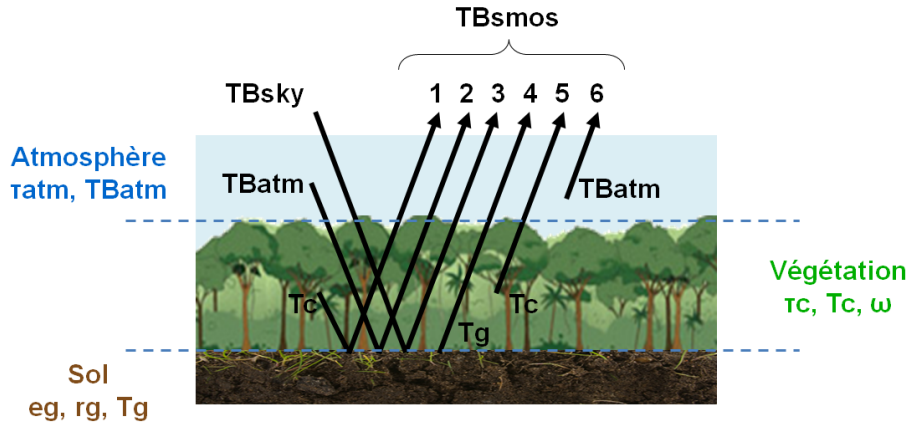


FIGURE 6.1 – Schéma des contributions au signal observé par le radiomètre en orbite avec trois couches : le sol (émissivité e_g , réflectivité r_g , température effective du sol T_g), la végétation (épaisseur optique τ , température T_c , albedo ω) et l'atmosphère (épaisseur optique de l'atmosphère τ , émission de l'atmosphère T_{Batm} , émission cosmique T_{Bsky}).

cela, plusieurs orbites ont été sélectionnées en Amérique du Nord et en Australie pour tester la distribution spatiale de l'humidité alors que les quatre sites test aux États-Unis ont été utilisés pour observer l'aspect temporel.

L'humidité est la variable que l'on souhaite observer et l'épaisseur optique de la végétation (qui correspond au ratio du signal qui a été absorbé par la végétation) est très souvent inconnu. Les algorithmes SMOS et LPRM permettent de retrouver ce paramètre. Sous certaines conditions, SMOS utilise une méthode d'optimisation afin de retrouver à la fois l'humidité et l'épaisseur optique. LPRM utilise le rapport de polarisation pour estimer l'épaisseur optique et une optimisation permet de retrouver l'humidité du sol. Le terme d'optimisation définit ici un processus itératif qui consiste à trouver des valeurs de paramètres telles que la différence entre le signal observé et le modèle de l'algorithme soit la plus faible possible. Il sera donc également possible de comparer les épaisseurs optiques retrouvées avec l'algorithme LPRM adapté aux acquisitions SMOS avec les épaisseurs optiques originales de SMOS et d'AMSR-E.

6.2 Modèle de transfert radiatif

6.2.1 Modèle de transfert radiatif à base de couches

Dans cette partie est présenté le modèle qui permet de relier le signal acquis par un radiomètre embarqué sur satellite et certains paramètres physiques du sol, de la végétation et de l'atmosphère (fig. 6.1). La théorie de transfert radiatif est présentée plus en détails dans Ulaby et al. (1986).

6.2.1.1 Modèle de transfert radiatif à trois couches

Un satellite en orbite quasi-polaire se situe à une altitude d'environ 800 km. Les satellites comme SMOS et AMSR-E embarquent à leur bord des radiomètres qui captent de manière passive le rayonnement micro-onde naturel qu'émet la Terre. Puisque tout objet qui possède une température supérieure à 0 K émet des ondes dans le domaine des micro-ondes, le sol, la végétation et l'atmosphère contribuent au signal observé par les radiomètres. Lorsqu'un

capteur acquiert un signal, il est modélisé par une somme de multiples contributions qui sont généralement regroupées comme suit (fig. 6.1) :

- atmosphère (6 sur la fig. 6.1),
- végétation passant à travers l'atmosphère (5),
- sol passant à travers la végétation puis l'atmosphère (4),
- végétation réfléchi par le sol puis traversant la végétation et l'atmosphère (1),
- atmosphère traversant la végétation, réfléchi par le sol puis traversant la végétation et l'atmosphère (2),
- rayonnement cosmique traversant l'atmosphère et la végétation puis réfléchi par le sol et repassant à travers la végétation et l'atmosphère (3).

La réflexion sur la végétation est très souvent négligée ainsi que les réflexions multiples dans la canopée (ATBD, 2010).

Plusieurs paramètres permettent de retranscrire ces phases de traversée et de rebond :

- atmosphère : τ_{atm} est l'épaisseur optique de l'atmosphère (coefficient d'atténuation),
- végétation : τ_c est l'épaisseur optique de la végétation (coefficient d'atténuation), ω est l'albedo (coefficient de diffusion), on définit la transmissivité de la végétation $\Gamma_c = \exp(-\tau_c)$,
- sol : e_g est le coefficient d'émissivité du sol, r_g est le coefficient de réflectivité du sol.

Il est ainsi possible de retranscrire ces contributions sous forme d'équations :

1. émission descendante de la couche de végétation

$$T_c \cdot (1 - \omega) \cdot (1 - \exp(-\tau_c)) \cdot r_g \cdot \exp(-\tau_c) \cdot \exp(-\tau_{atm}) \quad (6.1)$$

2. émission descendante de l'atmosphère

$$TB_{atm} \cdot \exp(-\tau_c) \cdot r_g \cdot \exp(-\tau_c) \cdot \exp(-\tau_{atm}) \quad (6.2)$$

3. émission galactique

$$TB_{sky} \cdot \exp(-\tau_{atm}) \cdot \exp(-\tau_c) \cdot r_g \cdot \exp(-\tau_c) \cdot \exp(-\tau_{atm}) \quad (6.3)$$

4. émission de la surface du sol

$$e_g \cdot T_g \cdot \exp(-\tau_c) \cdot \exp(-\tau_{atm}) \quad (6.4)$$

5. émission ascendante de la couche de végétation

$$T_c \cdot (1 - \omega) \cdot (1 - \exp(-\tau_c)) \cdot \exp(-\tau_{atm}) \quad (6.5)$$

6. émission ascendante de l'atmosphère

$$TB_{atm} \quad (6.6)$$

TB_{atm} est l'émission de l'atmosphère, TB_{sky} est le rayonnement galactique, T_c est la température de la végétation et T_g est la température effective du sol. T_g n'est pas la température de la surface du sol car une couche d'épaisseur variant avec la fréquence d'observation participe au signal observé par le radiomètre. La température dite effective tient compte de cette épaisseur ou profondeur de pénétration.

La température de brillance acquise par le radiomètre est la somme de ces contributions dont les équations précédentes sont des approximations. Cette température de brillance est bi-polarisée avec une polarisation horizontale (h) et une polarisation verticale (v). L'émissivité e_g ,

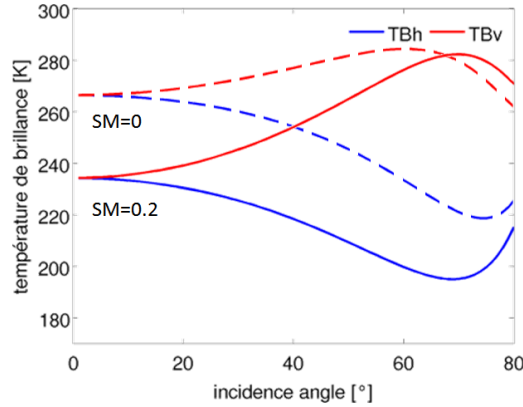


FIGURE 6.2 – Influence de l’humidité sur le profil angulaire des températures de brillance simulées en bande L (modèle LPRM, Owe et al. (2001)). L’humidité a tendance à diminuer les températures de brillance mais cette différence s’atténue avec l’angle d’observation.

la réflectivité r_g , l’albedo ω et l’épaisseur optique de la végétation τ_c sont également polarisés. Si l’indice p représente la polarisation alors la température de brillance peut s’écrire (diffusion à l’ordre 0) :

$$\begin{aligned}
 TB_{(p)} = & TB_{atm} + \exp(-\tau_{atm}) \cdot [TB_{atm} + TB_{sky} \cdot \exp(-\tau_{atm})] \cdot r_{g(p)} \cdot \exp(-2\tau_{c(p)}) \\
 & + \exp(-\tau_{atm}) \cdot [e_{g(p)} \cdot T_g \cdot \exp(-\tau_{c(p)}) \\
 & + T_c \cdot (1 - \omega_{(p)}) \cdot (1 - \exp(-\tau_{c(p)})) \cdot (1 + r_{g(p)} \cdot \exp(-\tau_{c(p)}))]
 \end{aligned} \quad (6.7)$$

Cette expression de la température de brillance, qui correspond au rayonnement acquis par un radiomètre sur orbite, est la base du modèle LPRM et de l’algorithme implémenté pour retrouver l’humidité SMOS. Cependant, SMOS présente plusieurs modèles de ce genre pour chacune de ses classes d’occupation des sols afin d’estimer leurs contributions au plus juste. Dans le modèle LPRM, la diversité du champ de vue n’est pas prise en compte et l’équation (6.7) est censée représenter l’ensemble de la surface observée. Il est important de remarquer que la température de brillance ainsi exprimée dans l’équation (6.7) dépend également et très fortement de l’angle d’observation θ .

6.2.1.2 Constante diélectrique du sol

La première étape est de calculer la constante diélectrique du sol qui est une propriété physique décrivant la réponse de la surface observée à un champ électrique. Les technologies micro-ondes se basent alors sur le large contraste qui existe entre la constante diélectrique de l’eau ($\varepsilon \sim 80$) et celle d’un sol sec ($\varepsilon \sim 4$) à 1.4 GHz. Les propriétés diélectriques d’un mélange sol-eau ($\varepsilon \sim 4$ -40) impactent l’émission micro-onde naturelle de ce sol (Schmugge, 1985).

La constante diélectrique est un nombre complexe. Sa partie réelle ε' est déterminée par les caractéristiques de propagation de l’énergie dans le sol, alors que sa partie imaginaire ε'' représente les pertes d’énergie. Le sol étant un mélange d’air, d’eau et de différentes particules, sa constante diélectrique est une combinaison des constantes diélectriques de chacun de ses composants. Elle est influencée par la texture du sol, sa porosité, sa salinité, sa température (très peu) et la fréquence d’observation. Il existe plusieurs modèles de constante diélectrique et parmi eux, on trouve le modèle de Wang-Schmugge (Wang & Schmugge, 1980), de Dobson (Dobson et al., 1985) ou de Mironov (Mironov et al., 2004). Ces trois modèles de constante

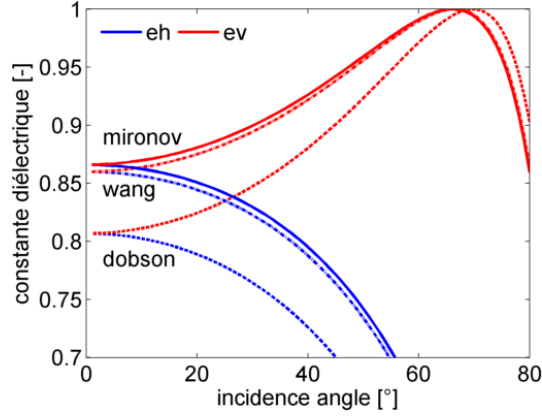


FIGURE 6.3 – Différence de modélisation de la constante diélectrique pour un sol d'humidité 0.15 m³/m³ composé de 10% de sable et 50% d'argile en utilisant les modèles de Wang-Schmugge, Dobson et Mironov.

diélectrique sont décrits en annexe C et un exemple de différence de modélisation est présenté en figure 6.3.

L'humidité du sol influe beaucoup sur la constante diélectrique et donc sur les températures de brillance acquises par un radiomètre. La figure 6.2 montre l'impact de l'humidité du sol sur le profil angulaire de la température de brillance simulée en polarisation verticale et horizontale. Un sol humide aura une température de brillance en générale plus faible qu'un sol sec. Cependant cette différence est moins visible dans les grands angles en polarisation verticale. De plus la valeur maximale de la température de brillance verticale se trouve pour un angle plus élevé alors que la valeur minimale de la polarisation horizontale se situe à un angle plus petit.

6.2.1.3 Réflectivité du sol

L'humidité de surface influe sur l'émissivité du sol et également sur sa réflectivité. Les lois de Fresnel (Ulaby et al., 1986) permettent de calculer la réflectivité d'une surface lisse r en fonction de l'angle d'observation θ et de la constante diélectrique du sol k ou ε :

$$r_H(\theta) = \left| \frac{\cos \theta - \sqrt{k - \sin^2 \theta}}{\cos \theta + \sqrt{k - \sin^2 \theta}} \right|^2 \quad (6.8)$$

$$r_V(\theta) = \left| \frac{k \cos \theta - \sqrt{k - \sin^2 \theta}}{k \cos \theta + \sqrt{k - \sin^2 \theta}} \right|^2 \quad (6.9)$$

Ces équations permettent de caractériser un sol lisse. En réalité, la surface observée est rugueuse et il faut prendre en compte la rugosité du sol.

6.2.1.4 Rugosité

La rugosité de surface augmente l'émissivité due à la diffusion d'une plus grande surface d'émission (Schmugge, 1985). La rugosité diminue également la sensibilité de l'émissivité aux variations de l'humidité du sol, réduisant ainsi l'écart entre l'émissivité d'un sol sec et d'un sol mouillé (Wigneron et al., 2011). Il existe plusieurs modèles reliant l'émissivité e et la rugosité h , par exemple celui de Choudhury et al. (1979) qui est décrit par :

$$e_{(p)}(\theta) = 1 - r_{(p)} \cdot \exp(-h \cdot \cos^2 \theta) \quad (6.10)$$

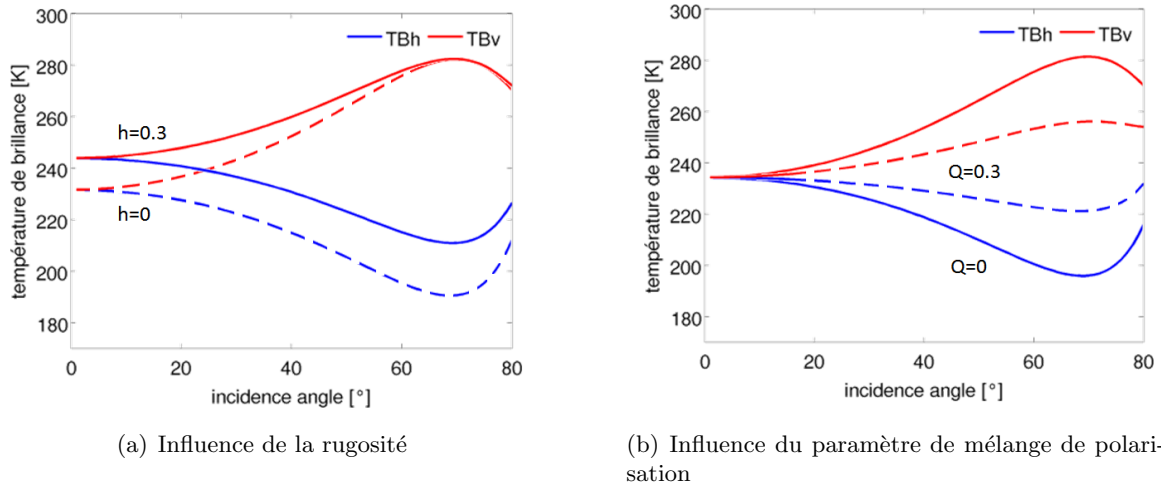


FIGURE 6.4 – Influence de la rugosité et du paramètre de mélange de polarisation sur le profil angulaire de la température de brillance simulée en bande L (modèle LPRM, Owe et al. (2001)). Une augmentation de la rugosité aura tendance à augmenter la température de brillance. Le paramètre de mélange de polarisations rapproche les polarisations verticales et horizontales de la température de brillance.

avec h le paramètre de rugosité qui est relié à la variation de hauteur quadratique moyenne et à la longueur de corrélation; θ l'angle d'incidence d'observation; r la réflectivité d'un sol lisse calculé avec les équations (6.9); p la polarisation (V ou H). De nombreux modèles ont été développés pour la rugosité de surface incluant des paramètres de mélange de polarisations (Wigneron et al., 1993a, Wang & Choudhury, 1981, Escorihuela et al., 2007) :

$$e_{(p)}(\theta) = 1 - ((1 - Q) \cdot r_{(p)} + Q \cdot r_{(\bar{p})}) \cdot \exp(-h(SM) \cdot \cos^{N_p} \theta) \quad (6.11)$$

avec p la polarisation (V ou H), \bar{p} la polarisation complémentaire (H ou V), Q le paramètre de mélange ou de couplage des polarisations et N_p le paramètre d'ajustement de la dépendance de la rugosité à l'angle θ .

Des valeurs moyennes de rugosité représentant l'intégralité d'un champ de vue d'un satellite (40-50 km pour SMOS ou AMSR-E) sont clairement difficiles à estimer. De plus la majorité des études menées sur la rugosité a été menée à une échelle locale avec des radiomètres installés sur site (Escorihuela et al., 2007, Wigneron et al., 2011).

Les paramètres de rugosité et de mélange influent sur la température de brillance acquise par un radiomètre embarqué sur satellite. La figure 6.4 montre leurs influences sur le profil angulaire de température de brillance en bande L. Une augmentation de la rugosité h fait augmenter la surface d'émission du sol, ce qui se traduit par une élévation de la température de brillance aux faibles angles. Cette différence s'estompe en polarisation verticale dans les grands angles alors qu'elle est toujours visible en polarisation horizontale. Le paramètre de mélange de polarisation Q rapproche les polarisations verticale et horizontale.

6.2.1.5 Effets de la végétation

Les effets de la végétation sur l'émission micro-onde telle qu'elle est mesurée par satellite sont doubles : (1) la végétation absorbe ou diffuse les radiations issues du sol et (2) la végétation émet également sa propre énergie micro-onde. Sous une végétation très dense de type équatoriale, le rayonnement du sol se retrouve fortement atténué et le signal acquis par le radiomètre proviendra

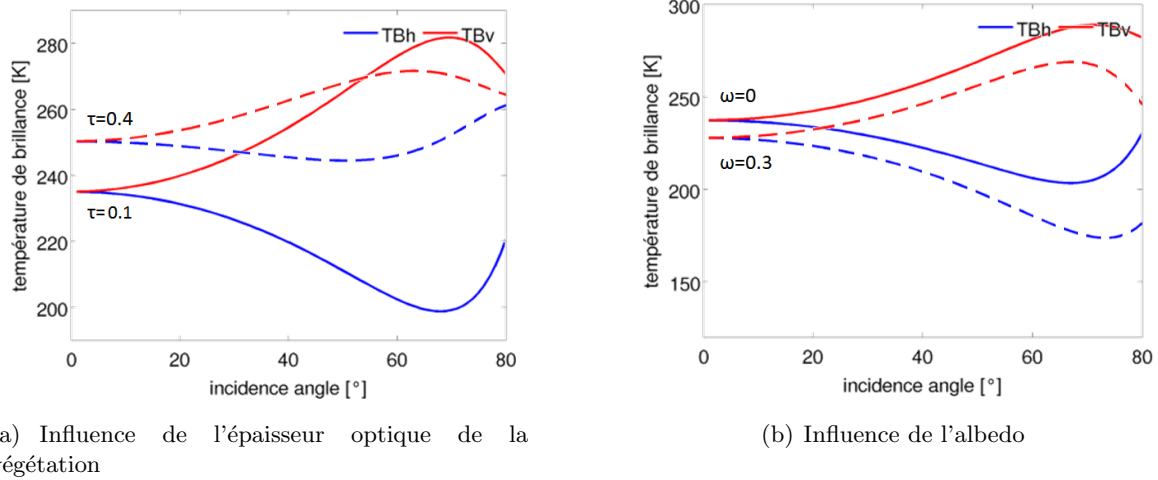


FIGURE 6.5 – Influence de la végétation sur le profil angulaire de la température de brillance simulée en bande L (modèle LPRM, Owe et al. (2001)). Plus l'épaisseur optique τ est élevée, plus la température de brillance au nadir est élevée et plus la différence de polarisation diminue. L'albédo abaisse le niveau moyen des températures de brillance, comme une translation verticale de tout le profil.

essentiellement de la végétation et non du sol (Ferrazzoli & Guerriero, 1996). L'ampleur de cette absorption du signal par la végétation dépend de sa composition, de son contenu en eau ou encore de la fréquence d'observation (fig. 2.3, p. 16). La bande L (1.4 GHz) est la bande de fréquence la plus adaptée pour retrouver l'humidité des sols mais les bandes C et X (6 et 10 GHz) ont été utilisées dans le passé pour retrouver cette variable. Parce que ces bandes sont sujettes à l'absorption et à la diffusion de la couche de végétation, il faut retirer au signal acquis la contribution de la végétation pour pouvoir retrouver l'humidité.

Plusieurs modèles ont été développés pour prendre en compte l'influence de la végétation avec des modèles linéaires empiriques (Choudhury et al., 1987, Ahmed, 1995) et des modèles plus physiques (Mo et al., 1982, Jackson et al., 1982, Jackson & O'Neill, 1990, Kerr & Njoku, 1990, Owe et al., 1992, Wigneron et al., 1995, Njoku & Li, 1999). L'épaisseur optique de la végétation τ_c est liée à sa densité et pour des fréquences d'observation inférieures à 10 GHz, elle peut s'exprimer comme une fonction linéaire de son contenu en eau (Jackson et al., 1982, Kerr & Njoku, 1990, Wigneron et al., 1993b). Dans la littérature, on peut trouver des valeurs de τ_c qui sont inférieures à 1.3 en bande C et à 0.4 en bande L pour différents couverts végétaux avec un contenu en eau stable autour de 3 kg/m³ (Mo et al., 1982, Pampaloni & Paloscia, 1986). Des calculs théoriques montrent que la sensibilité du signal aux variations de l'émissivité du sol diminue lorsque l'épaisseur optique de la végétation augmente (Schmugge, 1985).

L'albédo ω décrit la diffusion par la végétation. Il est fonction de la géométrie des plantes (orientation) et peut être calculé analytiquement (Wegmuller et al., 1995). Cependant des mesures expérimentales ont trouvé que pour certaines cultures, l'albédo variait entre 0.04 et 0.13 (Kirdiashev et al., 1979, Mo et al., 1982, Brunfeldt & Ulaby, 1984, Pampaloni & Paloscia, 1986, Becker & Choudhury, 1988, Van de Griend & Owe, 1994, Wigneron et al., 1995, Wegmuller et al., 1995, Kerr et al., 2012). Hormis une estimation de 0.05 sur une région semi-aride d'Afrique (Becker & Choudhury, 1988), peu de valeurs ont été retrouvées sur des espaces naturels.

Il existe des preuves expérimentales que l'épaisseur optique de la végétation et l'albédo sont polarisés. Cependant ces différences ne sont observables que sur des éléments de végétation qui manifestent une orientation préférentielle telles que les tiges verticales des herbes hautes

ou du maïs (Kirdiashev et al., 1979, Wigneron et al., 1995, Wegmuller et al., 1995). Au nadir (angle d'incidence nul), ces tiges ne sont plus visibles mais lorsque l'angle d'incidence augmente, l'orientation des tiges est de plus en plus visible, ce qui entraîne un effet plus important sur les émissions verticales. Mais à l'échelle d'un champ de vue d'un satellite avec une résolution de quelques dizaines de kilomètres, la végétation n'a pas d'orientation privilégiée. De ce fait, l'épaisseur optique et l'albédo sont supposés non polarisés : $\tau_{cv} = \tau_{ch}$ et $\omega_v = \omega_h$.

Ces deux paramètres de la végétation influent sur le profil angulaire de la température de brillance. La figure 6.5 montre l'influence de l'épaisseur optique et de l'albédo sur la température de brillance en bande L. L'épaisseur optique a tendance à augmenter le niveau des températures de brillance tout en rapprochant les deux polarisations (Prigent et al., 1997). L'effet de τ est surtout visible pour la polarisation horizontale qui devient indépendante de l'angle. L'albédo ω fait baisser la température de brillance, comme si une translation verticale avait été appliquée à l'ensemble du profil angulaire.

6.2.1.6 Température effective

Il a déjà été démontré que la température de brillance à 37 GHz en polarisation verticale est hautement corrélée à la température de surface (Basist et al., 1998, Owe & Van de Griend, 2001). Cette approche a été validée dans (Owe et al., 2001) en comparant les températures de surface modélisées avec les acquisitions du satellite TRMM (Tropical Rainfall Mapping Mission) et celles mesurées sur le réseau d'observation d'Oklahoma Mesonet aux États-Unis (Brock et al., 1995).

Les mesures de température de brillance quantifient l'intensité du rayonnement micro-onde de la surface observée. Avec l'approximation de Rayleigh-Jeans dans le domaine micro-onde, l'énergie émise par le sol est proportionnelle à la température thermodynamique. La température de brillance du sol peut ainsi s'exprimer comme une fonction de l'émissivité e et de la température effective du sol T_{eff} :

$$TB_{sol(p)} = e_{(p)} \cdot T_{eff} \quad (6.12)$$

Toute la couche du sol contribue à l'émission thermique du sol. L'intensité nette à la surface du sol est une superposition des intensités émises à différentes profondeurs du sol (Choudhury et al., 1982). La formulation théorique de la température effective demande une connaissance du profil vertical de la température du sol. Pour estimer T_{eff} avec le minimum d'information sur le profil du sol, de nombreuses paramétrisations ont été développées à partir : de l'infrarouge thermique (Wigneron et al., 1993b), de la température de l'air (Van de Griend & Owe, 1994), de la température de brillance à 37 GHz (Owe et al., 2001). Mais cette approximation est seulement valide à des fréquences supérieures à 10 GHz alors qu'aux fréquences moins élevées, la profondeur de pénétration est plus grande et des températures plus profondes influencent cette température effective (Choudhury et al., 1982).

Pour prendre en compte les effets du profil de la température du sol, Choudhury et al. (1982) ont proposé une paramétrisation de T_{eff} comme une fonction linéaire de la température du sol à deux profondeurs. Wigneron et al. (2001) proposent une amélioration en prenant en compte l'influence de l'humidité sur l'épaisseur de la couche de sol contribuant au signal : plus le sol est sec, plus profonde est la couche qui contribue. Holmes et al. (2006) ont proposé une nouvelle paramétrisation de T_{eff} basée sur la relation entre l'atténuation de l'énergie émise et la constante diélectrique de la surface du sol. Les différents modèles sont présentés en annexe D.

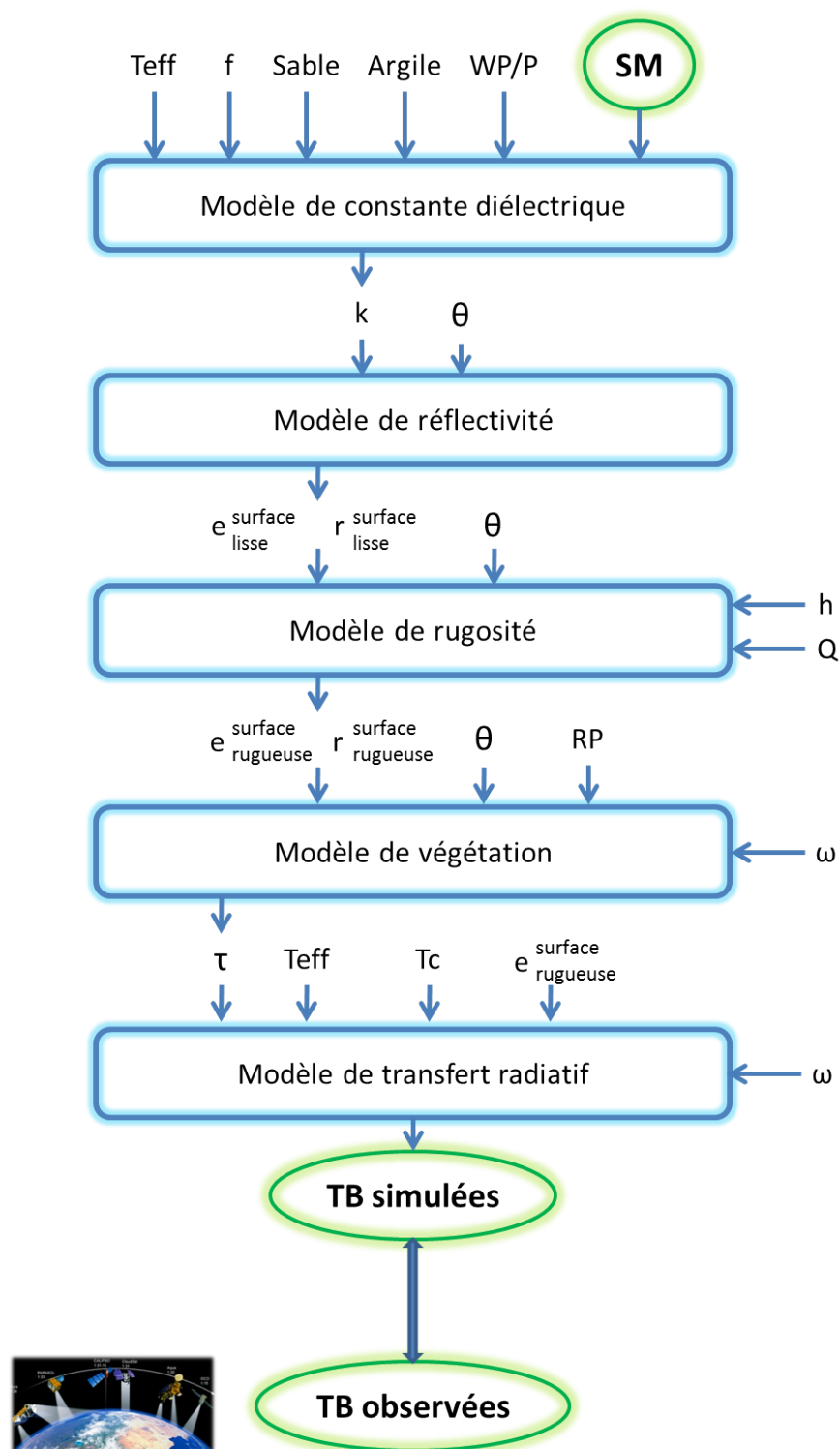


FIGURE 6.6 – Schéma de l'algorithme LPRM (*Land Parameter Retrieval Model*) en 5 phases : calcul de la constante diélectrique, calcul de la réflectivité du sol, prise en compte de la rugosité du sol, influence de la végétation puis modélisation du transfert radiatif afin de modéliser une température de brillance qui est ensuite comparée à la température acquise par satellite. L'humidité du sol retrouvée est celle qui minimise l'écart entre ces deux températures de brillance. Les flèches verticales représentent les entrées alors que les flèches verticales des paramètres.

6.2.2 Le modèle LPRM et les adaptations pour les acquisitions SMOS

Le modèle LPRM a été conçu pour retrouver l'humidité des sols en utilisant les températures de brillance acquises par les satellites AMSR-E, TRMM, SSM/I ou encore SMMR. Ces satellites observaient ou continuent d'observer la Terre avec un angle fixe à des fréquences en bande C, X ou encore Ka. Le schéma en figure 6.6 représente l'algorithme original LPRM. Il se compose de cinq modules qui ont été décrits précédemment : constante diélectrique, réflectivité de la surface lisse, rugosité pour surface rugueuse, végétation et enfin modèle de transfert radiatif. Ces cinq étapes permettent de modéliser une température de brillance qui est ensuite comparée à la température de brillance acquise par le radiomètre. L'humidité du sol est modifiée (entrée située en haut de la figure 6.6) jusqu'à ce que la différence entre les températures de brillance modélisées et observées soit à son minimum (différence pondérée par la précision des mesures) :

$$\min \left(\sum_i \left(\frac{TB_{Vobs}(\theta_i) - TB_{Vsimu}(\theta_i)}{RA_V(\theta_i)} \right)^2 + \sum_j \left(\frac{TB_{Hobs}(\theta_i) - TB_{Hsimu}(\theta_j)}{RA_H(\theta_j)} \right)^2 \right) \quad (6.13)$$

avec i, j le nombre d'acquisitions en polarisation horizontale et verticale respectivement ; RA la précision des mesures de température de brillance TB en K.

Les acquisitions du radiomètre embarqué sur le satellite SMOS sont différentes de celles acquises par AMSR-E par exemple. Ces acquisitions sont en bande L et sont multi-angulaires. Ces caractéristiques demandent une réelle adaptation de l'algorithme LPRM. de Jeu et al. (2009) ont mené une étude avec les données terrain de la campagne NAFE'05 réalisée en Australie en 2005 (Panciera et al., 2008). Ce jeu de données regroupe des mesures de l'humidité du sol à 5 cm de profondeur avec des observations aéroportées en bande L avec un angle d'incidence variant entre 2° et 44° .

Originellement, le modèle LPRM fixait le paramètre de rugosité h à 0.09 en bande C et à 0.18 en bande X. Escorihuela et al. (2007), Panciera et al. (2009) ont déjà montré que le paramètre de rugosité en bande L n'était pas constant et pouvait être relié à l'humidité. Sur le jeu de données NAFE'05, de Jeu et al. (2009) ont trouvé que la rugosité de surface avait une forte relation avec à la fois l'humidité et l'angle d'incidence. L'expression suivante avait été proposée pour leur terrain spécifique :

$$h = 0.4 - SM \cdot \theta^{1.5} \quad (6.14)$$

De plus, l'albédo de la végétation avait été fixée à 0 (en bande L, l'albédo est considérée comme très faible, Wigneron et al. (2004)) ainsi que le coefficient de mélange de polarisation :

$$\omega = 0 \quad (6.15)$$

$$Q = 0 \quad (6.16)$$

Pour l'adaptation du modèle LPRM à la bande L, ces propositions ont été gardées mais certains tests ont été réalisés en optimisant h et ω pour voir si ces modélisations réalisées sur des surfaces en Australie sont valides sur d'autres surfaces plus humides.

L'épaisseur optique de la végétation τ est retrouvée dans le LPRM en utilisant le rapport de polarisation (Meesters et al., 2005) :

$$RP(\theta) = \frac{TB_V(\theta) - TB_H(\theta)}{TB_V(\theta) + TB_H(\theta)} \quad (6.17)$$

Pour chaque angle observé, un rapport de polarisation est calculé et une valeur de τ est retrouvée. Les détails de ce calcul sont donnés en annexe E. Pour des températures de brillance

Tableau 6.1 – Modifications apportées au modèle LPRM pour son adaptation aux acquisitions multi-angulaires en bande L de SMOS.

LPRM	LPRM adapté à SMOS
mono-angulaire	multi-angulaire \Rightarrow optimisation des paramètres sur tout le profil angulaire ($\theta > 25^\circ$)
bandes C et X	bande L (paramétrisation proposée par de Jeu et al. (2009))
température effective du sol avec la température de brillance à 37 GHz	température effective du sol avec le modèle de Wigneron et al. (2001)
humidité <i>SM</i> optimisée	humidité <i>SM</i> optimisée

multi-angulaires, un rapport de polarisation est obtenu pour chaque angle. Malgré la division de l'épaisseur optique par le cosinus de l'angle d'observation, qui est censé enlever toute liaison avec l'angle d'incidence, la valeur obtenue n'est pas stable le long du profil (du simple au triple) certainement à cause des imperfections dans la mesure et des imperfections du modèle : étendue de la zone observée différente pour chaque angle (qui fait varier la proportion sol/végétation observée), interférences, etc. De ce fait, dans le modèle LPRM adapté à SMOS, le τ moyen a été pris en compte. Une autre version de ce modèle a été testé en optimisant l'épaisseur optique de la végétation. Les températures de brillance acquises par SMOS aux petits angles sont très bruitées et pour s'assurer que le rapport de polarisation est strictement positif, seules les températures de brillances acquises avec un angle supérieur à 25° sont utilisées.

LPRM utilise le modèle de Wang-Schmugge pour calculer la constante diélectrique. Le produit d'humidité SMOS utilisé ici (v400) a été retrouvé en utilisant le modèle de Dobson. Ces deux modèles ont été implémentés dans l'adaptation du LPRM à SMOS.

La température effective est originellement calculée à partir des températures de brillances à 37 GHz dans le LPRM. Puisque SMOS n'acquiert pas d'information à cette fréquence, la modélisation de la température par Wigneron et al. (2001) a été utilisée. Les températures de surface STL1 (*Soil Temperature Layer 1* : 0-7 cm) et les températures profondes STL3 (*Soil Temperature Layer 3* : 28 cm-1 m) du modèle ECMWF ont été utilisées pour calculer cette température effective. Les détails sont donnés en annexe D.

Les tableaux 6.1 et 6.2 résument les modifications apportées au modèle LPRM pour son adaptation aux acquisitions multi-angulaires en bande L de SMOS ainsi que les différentes options testées dans les prochains paragraphes : modèles A à H.

Tableau 6.2 – Options testées pour l’adaptation du LPRM aux acquisitions multi-angulaires en bande L de SMOS concernant le modèle de constante diélectrique, l’albédo et la rugosité (modèles A à H). RP pour rapport de polarisation, Eq. 6.16.

		LPRM original	A	B	C	D	E	F	G	H
Epaisseur optique de la végétation	RP	×	×			×	×	×	×	×
	RP lissé			×						
	RP itératif				×					
Modèle de constante diélectrique	Wang & Schmugge (1980)	×	×	×	×			×	×	×
	Dobson et al. (1985)				×			×	×	×
	Mironov et al. (2004)						×	×	×	×
Albédo	$\omega=0.06$	×								
	$\omega=0$		×	×	×	×	×		×	
	ω optimisé							×		×
Rugosité	$h=0.09, Q=0.115$ (bande C)	×								
	$h=0.18, Q=0.127$ (bande X)									
	$h=0, Q=0$		×	×	×	×	×	×		
	h optimisé, $Q=0$								×	
	$h(\text{SM}), Q=0$									×

6.3 Comparaison du modèle simplifié avec le modèle complexe SMOS

6.3.1 Paramétrisation du modèle (modèles A, B et C)

6.3.1.1 Profils angulaires de températures de brillance

La première étape est d'adapter simplement certains paramètres pour la bande L et de prendre en compte la multi-angularité des observations. L'humidité retrouvée est celle qui réduit au maximum la somme des écarts entre les températures de brillance modélisées et les températures de brillance observées.

La difficulté des modèles A, B et C se situe dans le calcul de l'épaisseur de la végétation. En effet, τ dépend du rapport de polarisation et de l'humidité SM, ce qui signifie qu'une valeur de τ est retrouvée pour chaque valeur de SM testée et pour chaque angle d'observation. Or ce modèle de transfert radiatif avait comme hypothèse que l'épaisseur optique n'avait pas de dépendance angulaire. De plus, le profil angulaire des températures de brillance modélisées n'est pas continu sur la gamme d'angles observés si τ varie avec l'angle (figs. 6.7 (a) et 6.8(a)).

Deux exemples d'application de cette première version du modèle sont donnés pour un profil de températures de brillance *propre* (fig. 6.7, où $TB_V > TB_H$) et pour un profil bruité plus difficile à interpréter (fig. 6.8). Ces deux profils de températures observées correspondent au même point mais à deux dates différentes. Les caractéristiques du sol sont supposées stables avec un pourcentage de sable de 40%, 20% d'argile avec une densité de 1.3 g/cm³.

A travers ces deux exemples, on observe que le profil de températures de brillance (TB) modélisé n'est pas régulier et est plus proche des observations (figs. 6.7(a), 6.8(a)). Cela est possible si la valeur du paramètre τ est différente pour chaque angle. L'épaisseur optique de la végétation sur les figures 6.7(e) et 6.8(e) varie en effet avec l'angle d'observation. Cependant, on peut noter que τ varie beaucoup plus dans le cas du profil plus bruité alors qu'on remarque une certaine stabilité dans le cas *propre* après 40° d'incidence. τ est différent à chaque angle d'incidence car ce paramètre dépend de l'inverse du rapport de polarisation (cf. annexe E).

Les rapports de polarisation sont montrés en figures 6.7(c) et 6.8(c). Quand le profil des TB mesurées est *propre*, le rapport de polarisation varie quasiment linéairement avec l'angle d'incidence. En revanche, il peut devenir chaotique et parfois négatif avec un profil plus bruité (fig. 6.8(c)), ce qui rend difficile le calcul de l'épaisseur optique de la végétation. Pour résoudre le problème de dépendance à l'angle ou de stabilité du paramètre τ , une régression linéaire du rapport de polarisation par rapport à l'angle est proposée (figs. 6.7(d) et 6.8(d)).

En effectuant une régression linéaire sur le rapport de polarisation, on effectue un lissage des températures de brillance (figs. 6.7(d) et 6.8(d)). Les TB modélisées sont maintenant plus régulières et lissées même si elles sont moins proches des TB observées (figs. 6.7(b) et 6.8(b)). L'épaisseur optique retrouvée à partir du profil *propre* est stable sur toute la gamme d'angles observés (fig. 6.7(f)). En revanche, la régression du rapport de polarisation ne suffit pas à la stabilisation de τ en cas de profil bruité (fig. 6.8(f)). Même si τ devient une fonction continue de l'angle, ce paramètre n'est pas stable et croît avec l'angle.

Une troisième possibilité d'application de ce modèle est de l'utiliser de manière itérative. La phase d'initialisation permet de retrouver une humidité optimisée avec τ moyen associé (moyenne arithmétique sur la gamme d'angles observés), comme réalisé précédemment, puis se poursuit par plusieurs itérations en fixant tour à tour l'humidité et l'épaisseur optique pour retrouver l'autre paramètre jusqu'à obtenir un couple de valeurs stables pour SM et τ . Puisque

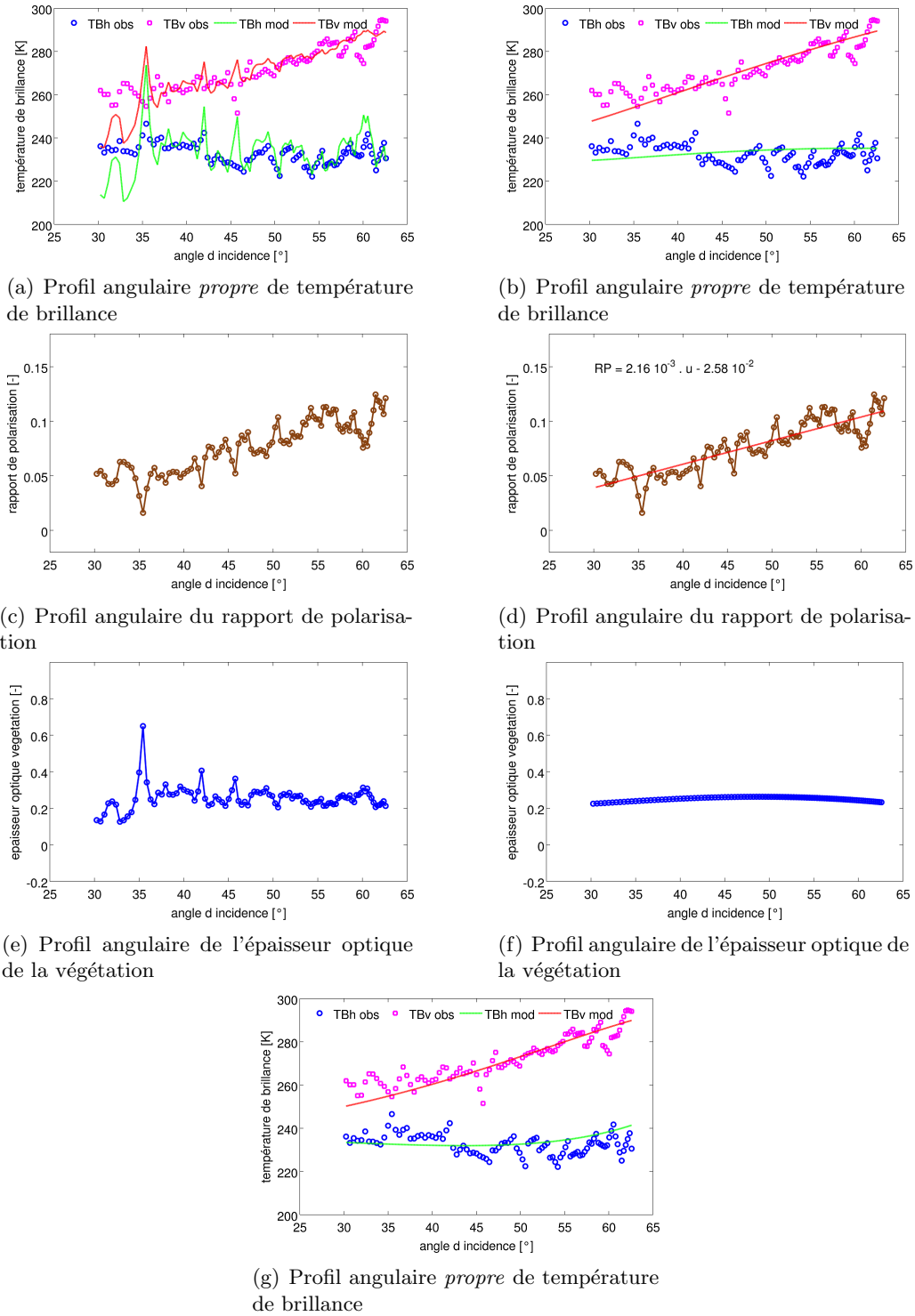


FIGURE 6.7 – Comportement du modèle LPRM sur un profil angulaire de température de brillance *propre*. La colonne de gauche (a,c,e) correspond au modèle original, la colonne de droite (b,d,f) au modèle avec le rapport de polarisation lissé. Le panneau du bas (g) correspond au modèle itératif. Les TB observées SMOS ainsi que les TB modélisées sont en (a,b,g), le rapport de polarisation par rapport à l'angle en (c,d) et l'épaisseur optique de la végétation pour chaque angle d'observation en (e,f).

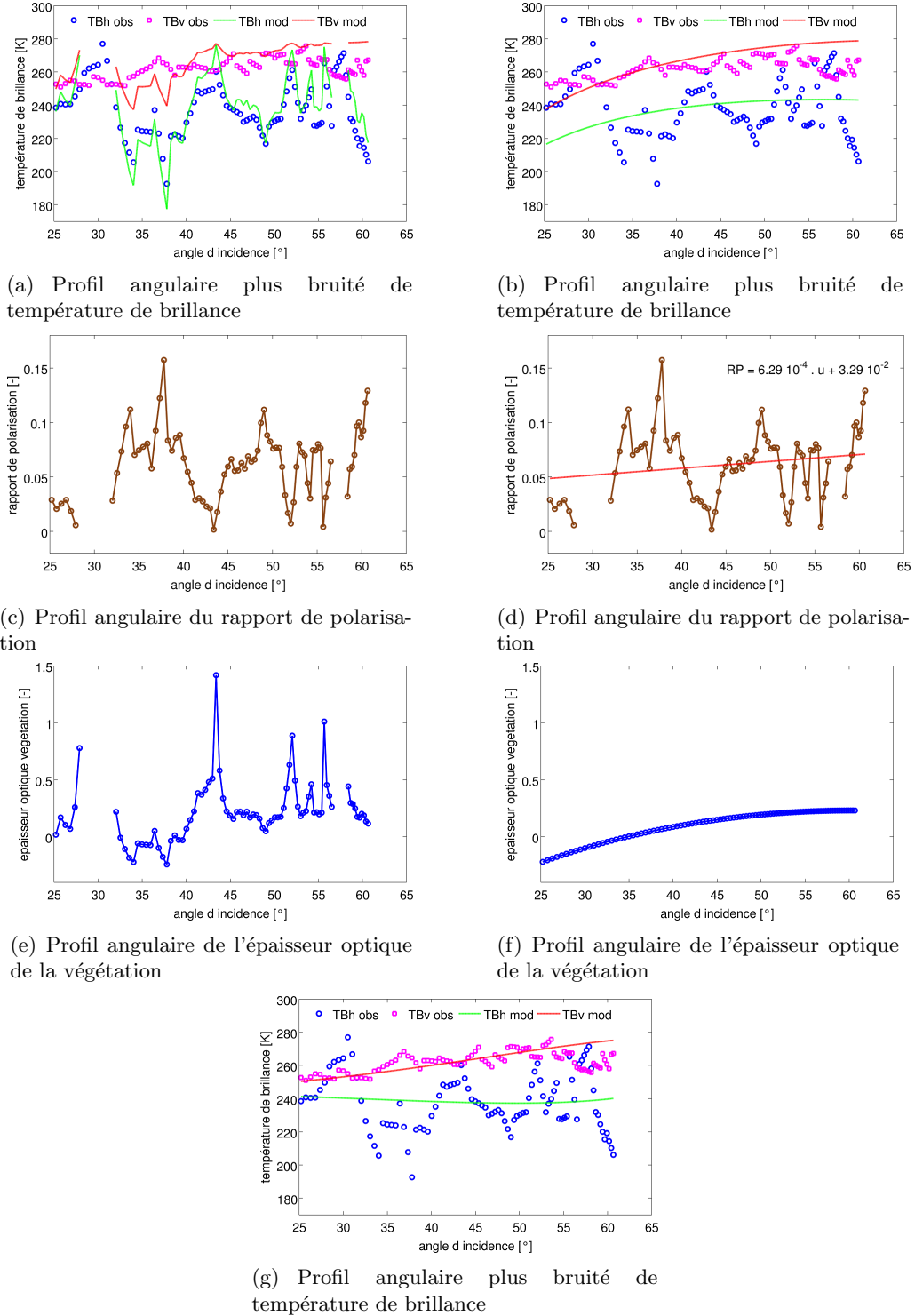


FIGURE 6.8 – Comportement du modèle LPRM sur un profil angulaire de température de brillance bruité. La colonne de gauche (a,c,e) correspond au modèle original, la colonne de droite (b,d,f) au modèle avec le rapport de polarisation lissé. Le panneau du bas (g) correspond au modèle itératif. Les TB observées SMOS ainsi que les TB modélisées sont en (a,b,g), le rapport de polarisation par rapport à l'angle en (c,d) et l'épaisseur optique de la végétation pour chaque angle d'observation en (e,f).

τ dépend de l'humidité SM, la décision d'arrêter le processus d'itérations se fait sur le paramètre d'humidité. Sur le profil *propre*, deux itérations ont été nécessaires (fig. 6.7(g)) et la différence avec l'application précédente (régression du rapport de polarisation) est essentiellement visible sur la modélisation des TB horizontales qui est dans ce cas plus courbée que la précédente. Sur le profil plus bruité, quatre itérations ont été nécessaires et de nombreuses différences sont à noter, notamment la courbure des TB modélisées.

Ces trois approches montrent des possibilités de modélisation différentes et représentent également trois couples de valeurs d'humidité et d'épaisseur optique. Pour les profils *propres* et bruités, voici les valeurs des paramètres retrouvés :

Modèle	Profil <i>propre</i>		Profil bruité	
Original	SM = 0.29	$\tau = 0.26$	SM = 0.10	$\tau = 0.22$
RP lissé	SM = 0.29	$\tau = 0.25$	SM = 0.08	$\tau = 0.11$
Itératif	SM = 0.30	$\tau = 0.26$	SM = 0.18	$\tau = 0.30$

A travers ces valeurs retrouvées, on remarque que pour le profil *propre*, l'approche choisie n'a que peu d'impact sur les résultats alors que pour le profil plus bruité, ce choix influence grandement les valeurs de SM et de τ (SM varie du simple au double). La première approche n'est pas correcte physiquement car le paramètre de végétation oscille avec l'angle d'observation. La seconde approche a tendance à trop lisser les températures de brillance dans le cas d'un profil plus bruité car le rapport de polarisation varie beaucoup. La troisième approche semble donc la plus correcte.

Cette section présentait deux exemples de profils angulaires de températures de brillance. Les deux prochaines sections exposent l'aspect spatial sur des parties d'orbite aux États-Unis et en Australie, puis l'aspect temporel sur deux bassins versants aux États-Unis : Little Washita et Walnut Gulch (présentés au chapitre 3).

6.3.1.2 Variation spatiale

L'aspect spatial est étudié dans cette section à travers deux exemples aux États-Unis et en Australie. Ces deux régions ont été sélectionnées car elles sont peu affectées par les interférences en bande L. Les dates d'étude ont été choisies de telle façon qu'un événement pluvieux soit visible : 16 Septembre 2010 pour les États-Unis et le 17 Septembre 2010 pour l'Australie.

Les trois approches du modèle LPRM ont été appliquées aux deux parties d'orbite sélectionnées : LPRM basique, LPRM avec le rapport de polarisation (RP) lissé et le LPRM itératif. Pour comparaison, les produits d'humidité SMOS sont également présentés (fig. 6.9).

Les distributions spatiales des trois approches sont très proches. La pluie au centre et à l'Est de la fauchée est correctement retranscrite par ces trois options. On observe également les zones de transition en vert entre les sols secs en jaune et les sols humides en bleu (fig. 6.9 (a), (b) et (c) aux États-Unis et (e), (f) et (g) en Australie). La différence la plus importante se situe au niveau du nombre de points où l'humidité est retrouvée et valide : les points où la valeur d'humidité retrouvée atteint une borne (0 ou $0.6 \text{ m}^3/\text{m}^3$) ne sont pas valides.

nb. de points	LPRM basique	régression du RP	LPRM itératif	Total
États-Unis	9104	6923	8433	9889
Australie	10051	8319	10068	10761

La méthode utilisant la régression du RP retrouve beaucoup moins de points avec environ 75% des points alors que le LPRM basique et itératif retrouvent une valeur d'humidité valide

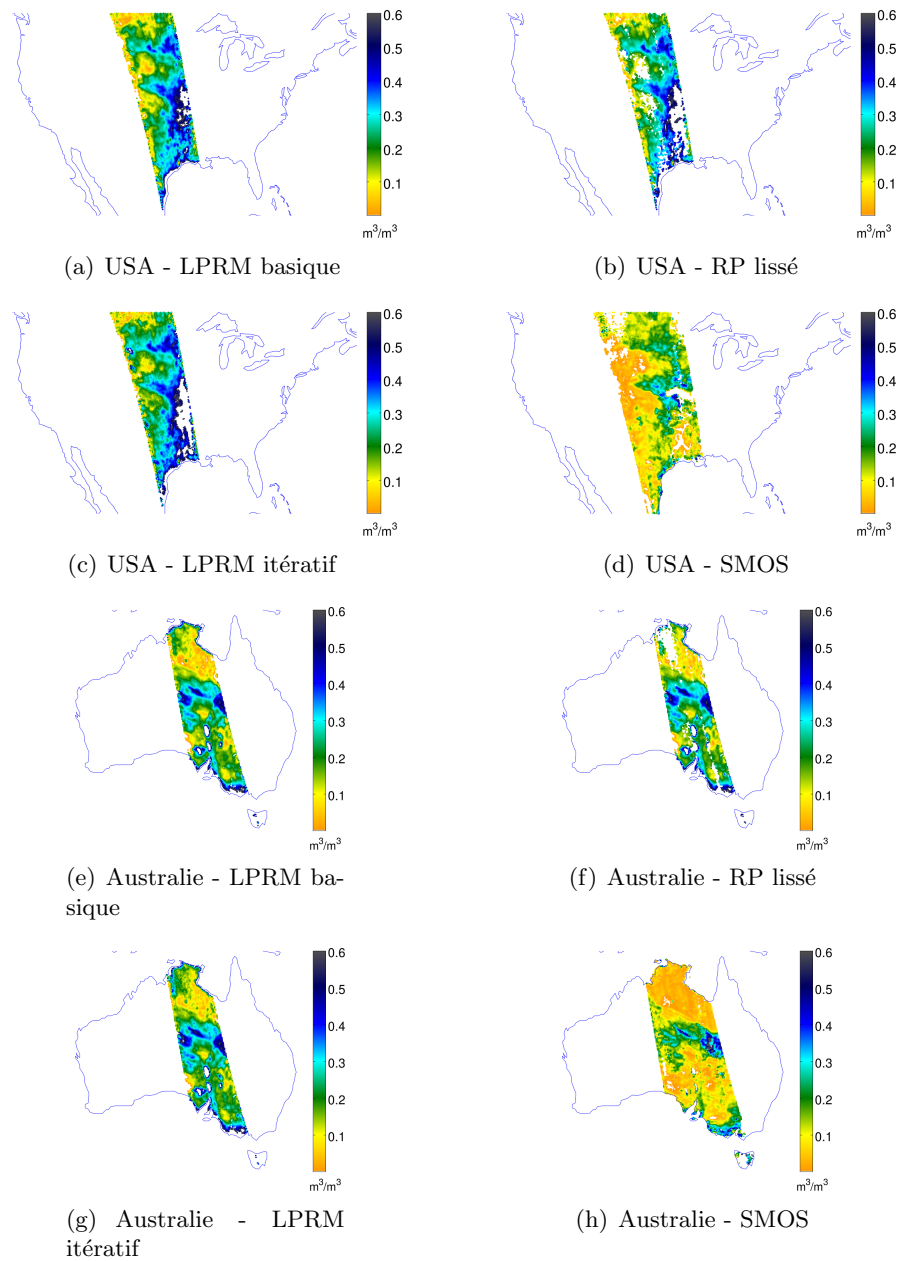


FIGURE 6.9 – Représentations spatiales des humidités retrouvées avec les trois approches LPRM avec celles de SMOS sur les Etats-Unis (16/09/2010) et sur l'Australie (16/09/2010) : (a,e) LPRM basique, (b,f) LPRM avec le rapport de polarisation (RP) lissé, (c,g) LPRM itératif et (d,h) SMOS.

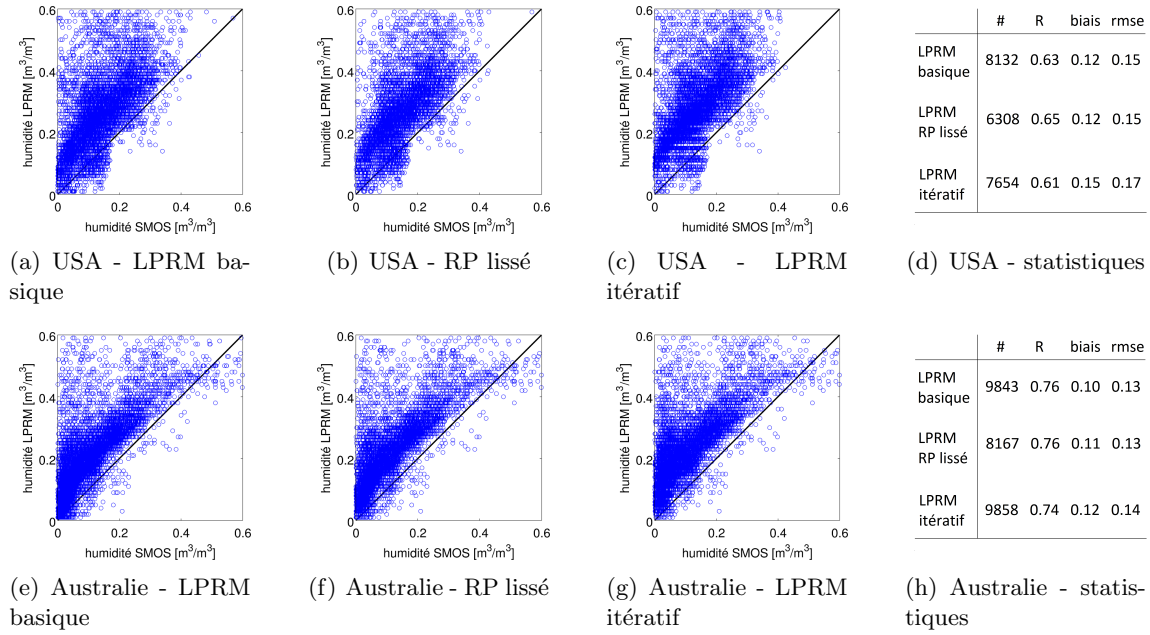


FIGURE 6.10 – Comparaison des trois approches d’application du modèle LPRM avec le produit d’humidité SMOS sur deux parties d’orbites aux États-Unis et en Australie : (a,d) LPRM basique, (b,e) LPRM avec le rapport de polarisation (RP) lissé et (c,f) LPRM itératif.

sur environ 90% des points. On peut également noter une augmentation des valeurs d’humidité retrouvées par le LPRM itératif par rapport aux deux autres méthodes (visible sur les États-Unis, fig. 6.9 (c)).

Les humidités retrouvées par le LPRM ont également été comparées au produit d’humidité SMOS (fig. 6.10). Les trois approches du modèle LPRM retrouvent des valeurs d’humidité plus élevées que le produit SMOS avec un biais moyen de $0.11 \text{ m}^3/\text{m}^3$ pour le LPRM basique, $0.11 \text{ m}^3/\text{m}^3$ avec le RP lissé et $0.13 \text{ m}^3/\text{m}^3$ pour le LPRM itératif. Les trois méthodes ont une corrélation équivalente avec SMOS, autour de 0.63 sur les États-Unis et de 0.75 sur l’Australie. On peut également noter la forte dispersion des nuages de points qui pourrait s’expliquer par la non-prise en compte de l’hétérogénéité de la surface par le LPRM contrairement au produit SMOS.

En conclusion, les distributions spatiales des trois approches sont très similaires et les valeurs d’humidité retrouvées sont toutes supérieures à celles retrouvées par SMOS. Les LPRM basique et itératif retrouvent une valeur d’humidité valide sur plus de points et le LPRM itératif a tendance à retrouver des valeurs plus élevées.

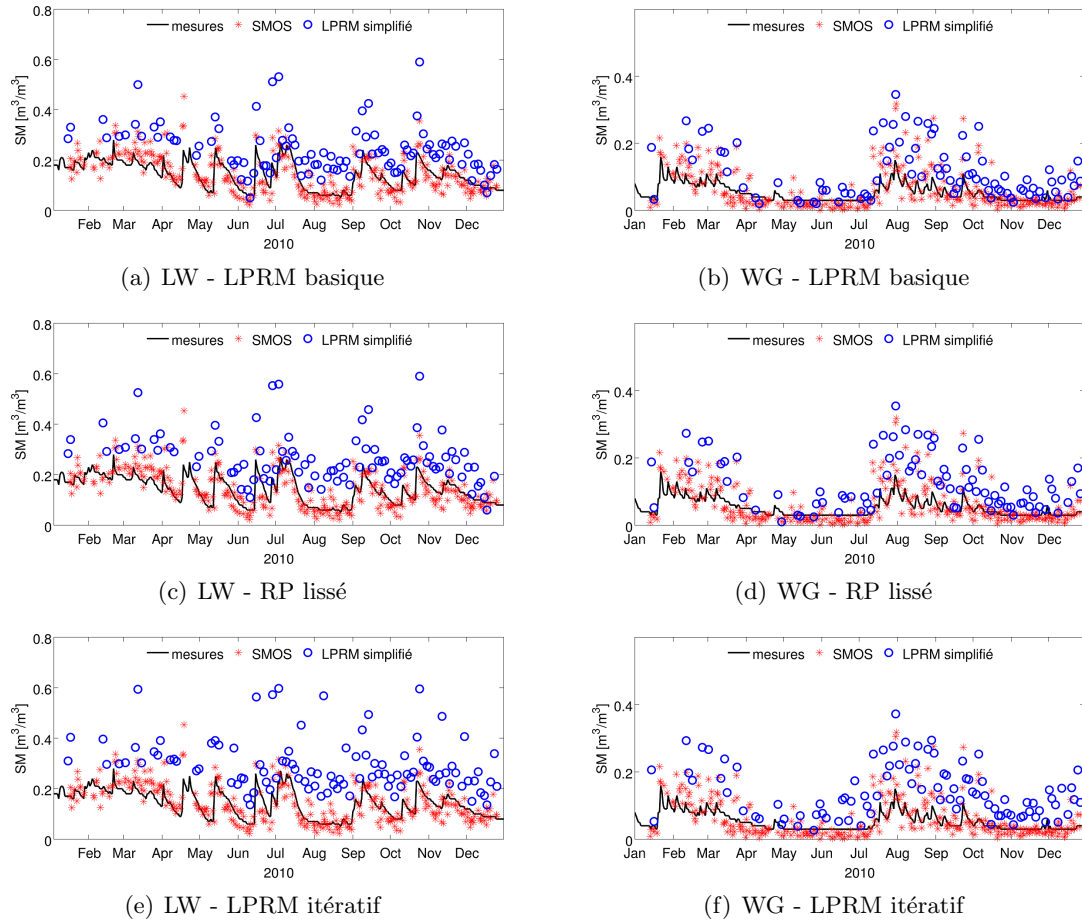


FIGURE 6.11 – Représentations temporelles des humidités retrouvées avec les trois approches du modèle LPRM sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) LPRM basique, (c,d) LPRM avec rapport de polarisation (RP) lissé et (e,f) LPRM itératif.

6.3.1.3 Variation temporelle

L'aspect temporel est étudié dans cette section à l'aide de deux bassins versants aux Etats-Unis tout au long de l'année 2010 : Little Washita, site subhumide avec de nombreux événements pluvieux et Walnut Gulch, site semi-aride avec très peu de pluies.

Les trois approches du modèle LPRM ont été appliquées aux deux séries temporelles : LPRM basique, LPRM avec le rapport de polarisation (RP) lissé et le LPRM itératif. Pour comparaison, les produits d'humidité SMOS et les humidités mesurées sur le terrain sont également présentés.

Les humidités retrouvées par les trois méthodes paraissent dispersées. En effet, sur Little Washita pendant la période sèche en Août, l'humidité mesurée est stable autour de $0.10 \text{ m}^3/\text{m}^3$ (fig. 6.11 (a), (c) et (e)). Les humidités retrouvées sont surestimées sur toute l'année et ne sont pas stables mais dispersées. Cependant, les valeurs du LPRM basique retranscrivent bien les événements pluvieux avec des valeurs hautes retrouvées suivies de phases de décroissance (pluies de Mai, Juin, Juillet par exemple). Les autres approches donnent également des valeurs très dispersées et la dynamique de l'humidité est plus difficile à reconnaître.

Sur Walnut Gulch, les humidités retrouvées sont également très dispersées (fig. 6.11 (b), (d) et (f)). Les faibles pluies en hiver puis en été sont bien captées mais sont largement surestimées. La période sèche d'Avril à Juillet n'est pas bien retranscrite puisque l'humidité retrouvée n'est pas stable. De plus, peu de points sont retrouvés sur cette période : 19 avec le LPRM basique,

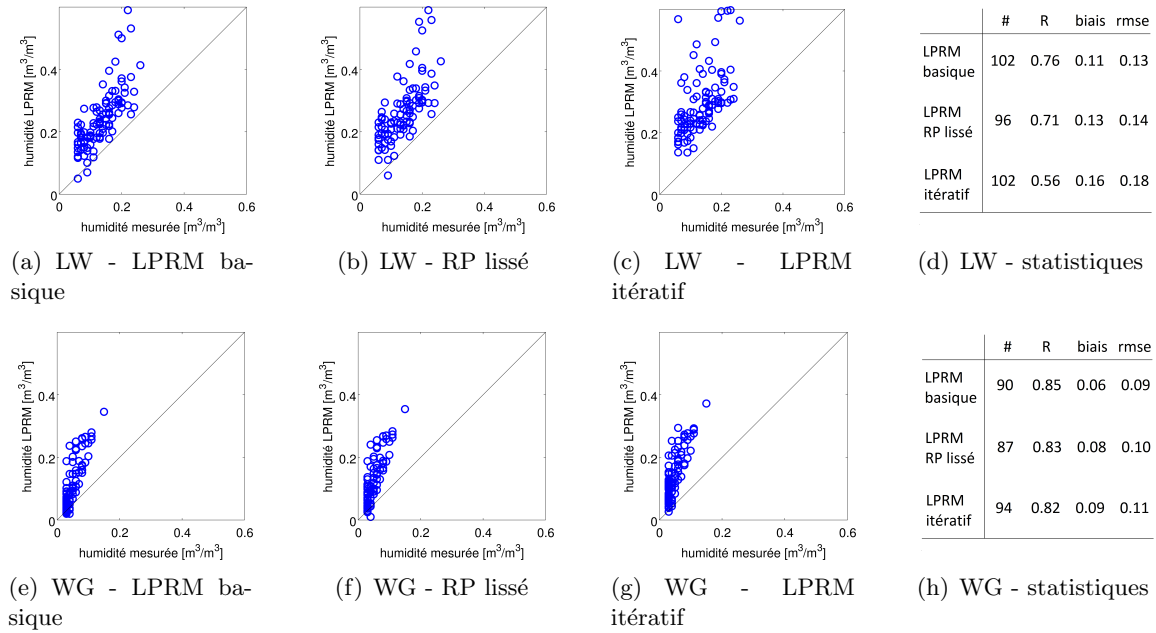


FIGURE 6.12 – Comparaison des humidités retrouvées par les trois approches du modèle LPRM avec les humidités mesurées sur Little Washita (LW) et Walnut Gulch (WG) : (a,e) LPRM basique, (b,f) LPRM avec rapport de polarisation (RP) lissé et (c,g) LPRM itératif. Les statistiques sont présentées en (d,h).

18 avec le LPRM avec le RP lissé et 23 avec le LPRM itératif. Cela peut être dû au modèle de constante diélectrique utilisé par le LPRM (Wang-Schmugge) qui n'est peut être pas adéquate pour les faibles valeurs d'humidité.

Les comparaisons avec les mesures effectuées sur le terrain (fig. 6.12) montrent bien les biais constatés sur les séries temporelles. Sur Little Washita, le biais est de $0.11 \text{ m}^3/\text{m}^3$ pour le LPRM basique, $0.13 \text{ m}^3/\text{m}^3$ pour le LPRM avec le RP lissé et $0.16 \text{ m}^3/\text{m}^3$ pour le LPRM itératif. Le niveau de corrélation est nettement plus faible pour le LPRM itératif ($R=0.56$) alors que le LPRM basique a une corrélation de 0.76 et le LPRM avec le RP lissé de 0.71.

Sur Walnut Gulch, le biais est d'une autre nature puisque les faibles valeurs d'humidité sont bien modélisées alors que de plus grands écarts existent pour les valeurs plus élevées. Avec un niveau de corrélation équivalent (autour de 0.83), le LPRM basique présente le plus faible biais avec $0.06 \text{ m}^3/\text{m}^3$ contre $0.08 \text{ m}^3/\text{m}^3$ pour le LPRM avec RP lissé et $0.09 \text{ m}^3/\text{m}^3$ pour le LPRM itératif.

En conclusion les distributions spatiales des trois approches du LPRM sur les deux fauchées testées sont très similaires. Sur l'aspect temporel à travers deux séries de mesures terrain, le LPRM basique semble fournir des humidités plus proches des mesures relevées sur le terrain. Cependant, les humidités retrouvées sont surestimées et plus spécialement durant les événements pluvieux. Cette surestimation peut provenir du modèle de constante diélectrique utilisé dans le LPRM.

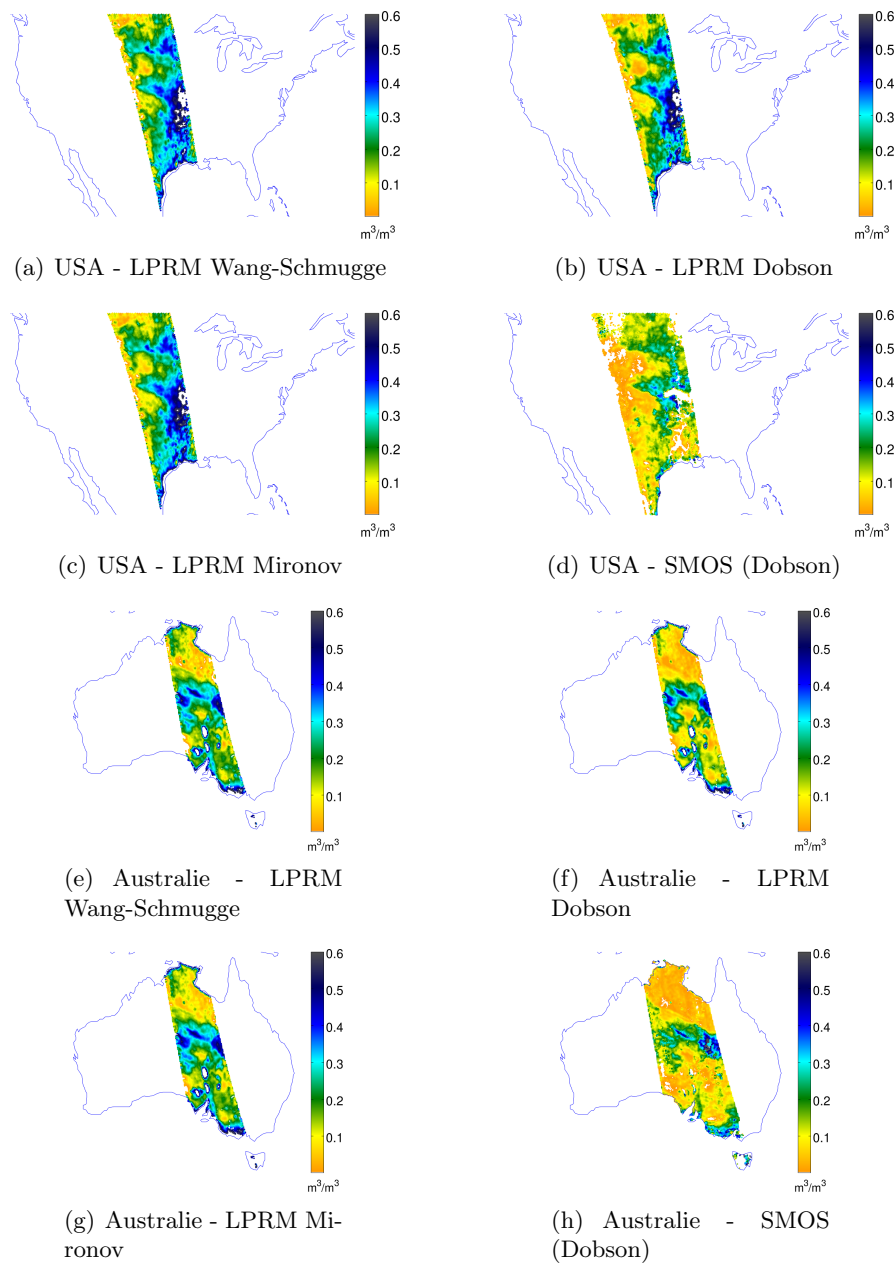


FIGURE 6.13 – Représentations spatiales des humidités retrouvées à partir des trois modèles de constante diélectrique avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,e) LPRM avec le modèle de Wang-Schmugge, (b,f) LPRM avec le modèle de Dobson, (c,g) LPRM avec le modèle de Mironov et (d,h) SMOS avec le modèle de Dobson.

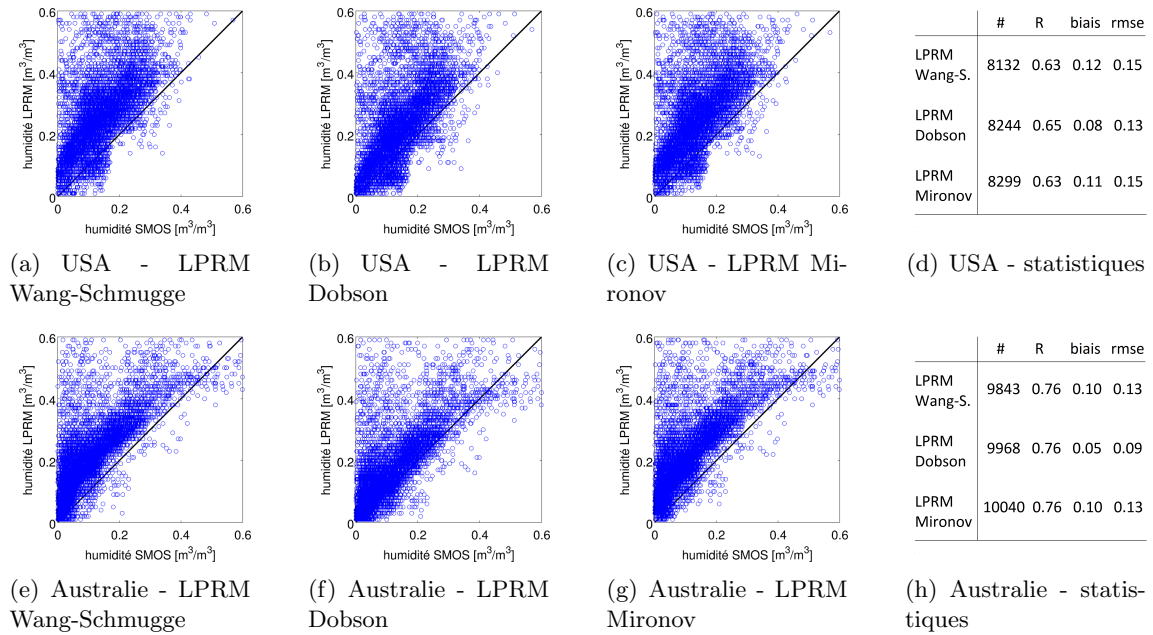


FIGURE 6.14 – Comparaison des humidités retrouvées avec le LPRM et les trois modèles de constante diélectrique avec celles de SMOS sur les Etats-Unis et sur l’Australie : (a,d) modèle de Wang-Schmugge, (b,e) modèle de Dobson et (c,f) modèle de Mironov.

6.3.2 Comparaison de trois modèles de constante diélectrique (modèles A, D et E)

Dans cette section, trois modèles de constante diélectrique sont étudiés : Wang-Schmugge (utilisé dans le LPRM, Wang & Schmugge (1980)), Dobson (utilisé pour le produit d’humidité SMOS, Dobson et al. (1985)) et Mironov (implémenté dans le prototype d’inversion des humidités SMOS, Mironov et al. (2004)). Ces trois modèles ont été testés avec le LPRM basique de la section précédente.

6.3.2.1 Variation spatiale

Les distributions spatiales sont encore une fois très similaires avec un nombre de points valides quasiment identique (environ 95% des points disponibles). Les points où l’humidité retrouvée n’est pas valide sont situés dans une zone très humide.

Aussi bien sur l’exemple des États-Unis que sur l’Australie, le modèle de constante diélectrique de Dobson assèche les valeurs d’humidité alors que le modèle de Mironov assèche un peu moins par rapport au modèle de Wang-Schmugge (fig. 6.13).

Le niveau de corrélation est équivalent pour les trois modèles de constante diélectrique, autour de 0.64 sur les États-Unis et de 0.76 en Australie. Les faibles valeurs d’humidité sont plutôt bien corrélées avec celles de SMOS, notamment en Australie alors que les hautes valeurs sont plus élevées avec le LPRM qu’avec SMOS, peu importe le modèle de constante diélectrique.

Avec un biais de $0.08 \text{ m}^3/\text{m}^3$ sur les Etats-Unis et de $0.05 \text{ m}^3/\text{m}^3$ sur l’Australie (fig. 6.14), le modèle de Dobson est celui qui produit des humidités les plus proches de celles de SMOS. Ce résultat est logique puisque l’humidité SMOS est produite avec le modèle de Dobson.

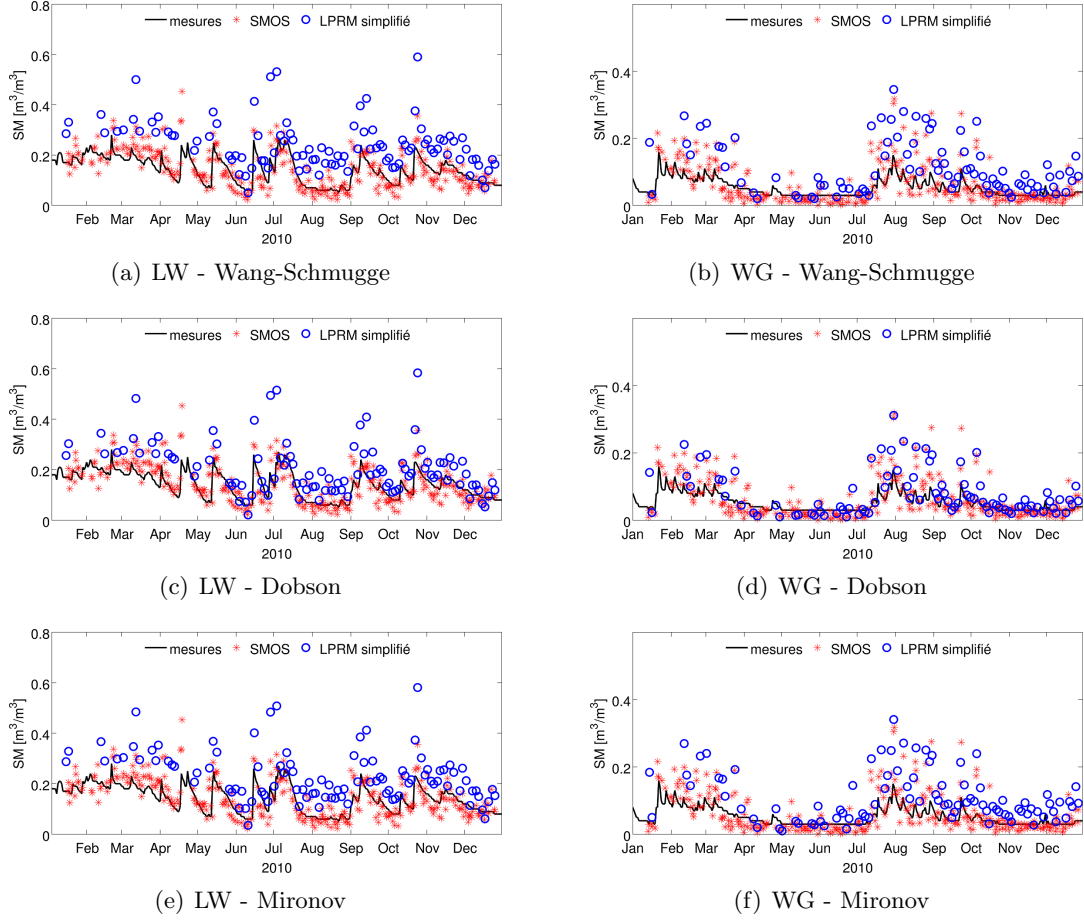


FIGURE 6.15 – Représentations temporelles des humidités retrouvées avec le LPRM et trois modèles de constante diélectrique sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.

6.3.2.2 Variation temporelle

Le modèle de constante diélectrique Wang-Schmugge est celui utilisé par le LPRM basique donc les séries temporelles présentées sont identiques à celles de la section précédente (fig. 6.15 (a) et (b)). Avec le modèle de Dobson (fig. 6.15 (c) et (d)), on observe une diminution des valeurs faibles de l'humidité, notamment sur la période sèche de Little Washita (LW) en Août et de Walnut Gulch (WG) en Avril-Juillet. En revanche, les surestimations lors des événements pluvieux sont encore présentes. Le modèle de Mironov se situe entre Wang-Schmugge et Dobson en terme de biais. Sur l'exemple de WG, Mironov modélise bien la période sèche d'Avril à Juillet mais surestime celle d'Octobre à Décembre.

Le niveau de corrélation est équivalent pour les trois modèles de constante diélectrique : autour de 0.77 pour LW et de 0.84 pour WG (fig. 6.16). En revanche, les biais et les erreurs sont différents puisque Dobson est plus proche des mesures terrain avec un biais de $0.07 \text{ m}^3/\text{m}^3$ sur LW (contre 0.09 et $0.11 \text{ m}^3/\text{m}^3$ pour Mironov et Wang-Schmugge) et de $0.03 \text{ m}^3/\text{m}^3$ sur WG (contre 0.06 et $0.06 \text{ m}^3/\text{m}^3$ pour Mironov et Wang-Schmugge). Comme précédemment, le biais est de nature différente sur LW et WG puisqu'il est constant sur LW alors que les faibles humidités modélisées sont très proches des mesures et s'éloignent dans les hautes valeurs sur WG.

La comparaison de ces trois modèles de constante diélectrique montrent que les distribution

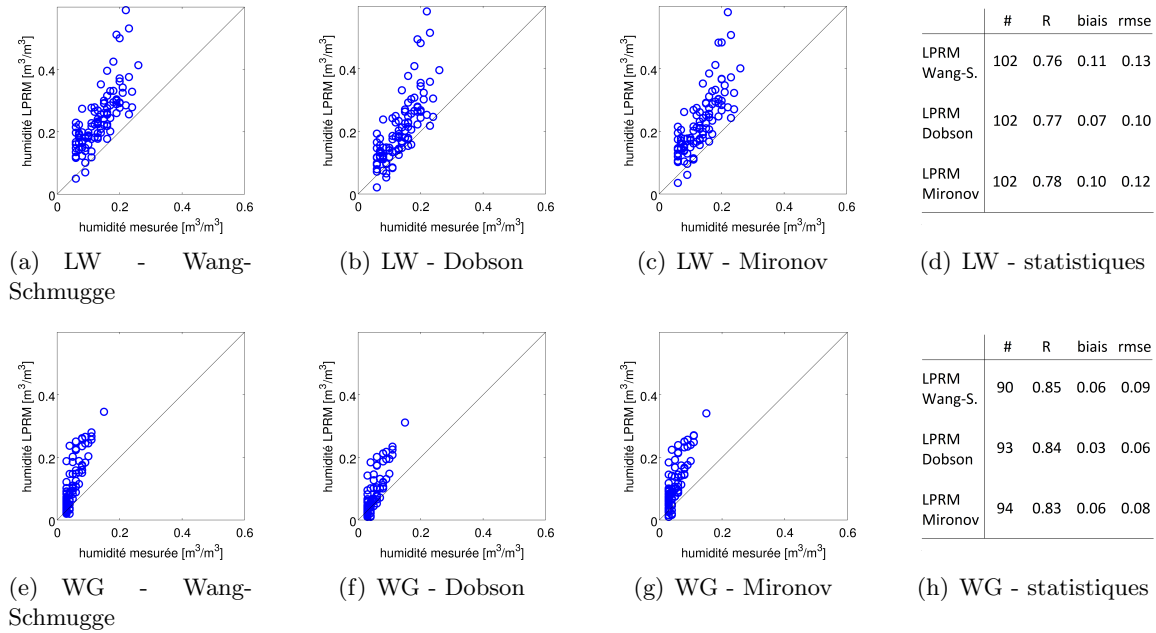


FIGURE 6.16 – Comparaison des humidités retrouvées par le modèle LPRM et trois modèles de constante diélectrique sur Little Washita (LW) et Walnut Gulch (WG) : (a,e) Wang-Schmugge, (b,f) Dobson et (c,g) Mironov. Les statistiques sont présentées en (d,h).

spatiales sont une nouvelle fois très proches. Cependant le modèle de constante diélectrique de Dobson se détache car il permet de modéliser des humidités beaucoup plus proches des mesures terrain. Les événements pluvieux sont tout de même encore surestimés peu importe le modèle de constante diélectrique utilisé.

Dans la prochaine section, d'autres options du modèle LPRM sont explorées afin de se rapprocher encore des mesures terrain.

6.3.3 Optimisation d'autres paramètres : albédo et rugosité (modèles F, G et H)

Lors de l'adaptation du LPRM aux observations multi-angulaires en bande L de SMOS, plusieurs hypothèses ont été posées. Concernant la végétation, il a été supposé qu'en bande L, il n'y avait aucune diffusion, i.e., $\omega = 0$. Il a été également supposé que la rugosité n'affectait pas le signal émis par la surface du sol, i.e., $h=0$. Dans cette section, ces deux hypothèses sont remises en question afin de mieux comprendre quels impacts ces deux paramètres de végétation et de rugosité ont réellement sur les températures de brillance modélisées et sur les valeurs d'humidité retrouvées.

6.3.3.1 Modèle F : albédo optimisé et rugosité nulle

Dans cette partie, le paramètre d'albédo ω n'est plus fixé à 0 mais est optimisé de la même manière que l'humidité du sol SM. Il s'agit donc de trouver le couple de valeurs (SM, ω) qui réduira au maximum l'écart entre les TB observées et les TB modélisées. La gamme de valeurs possibles pour SM est toujours de 0 à 0.6 m³/m³ et la gamme de ω de 0 à 1.

Les trois modèles de constante diélectrique (Wang-Schmugge, Dobson et Mironov vus dans la section précédente) sont conservés ici pour mieux comprendre l'impact de ces changements sur l'humidité retrouvée. La distribution spatiale de l'humidité est toujours étudiée sur deux

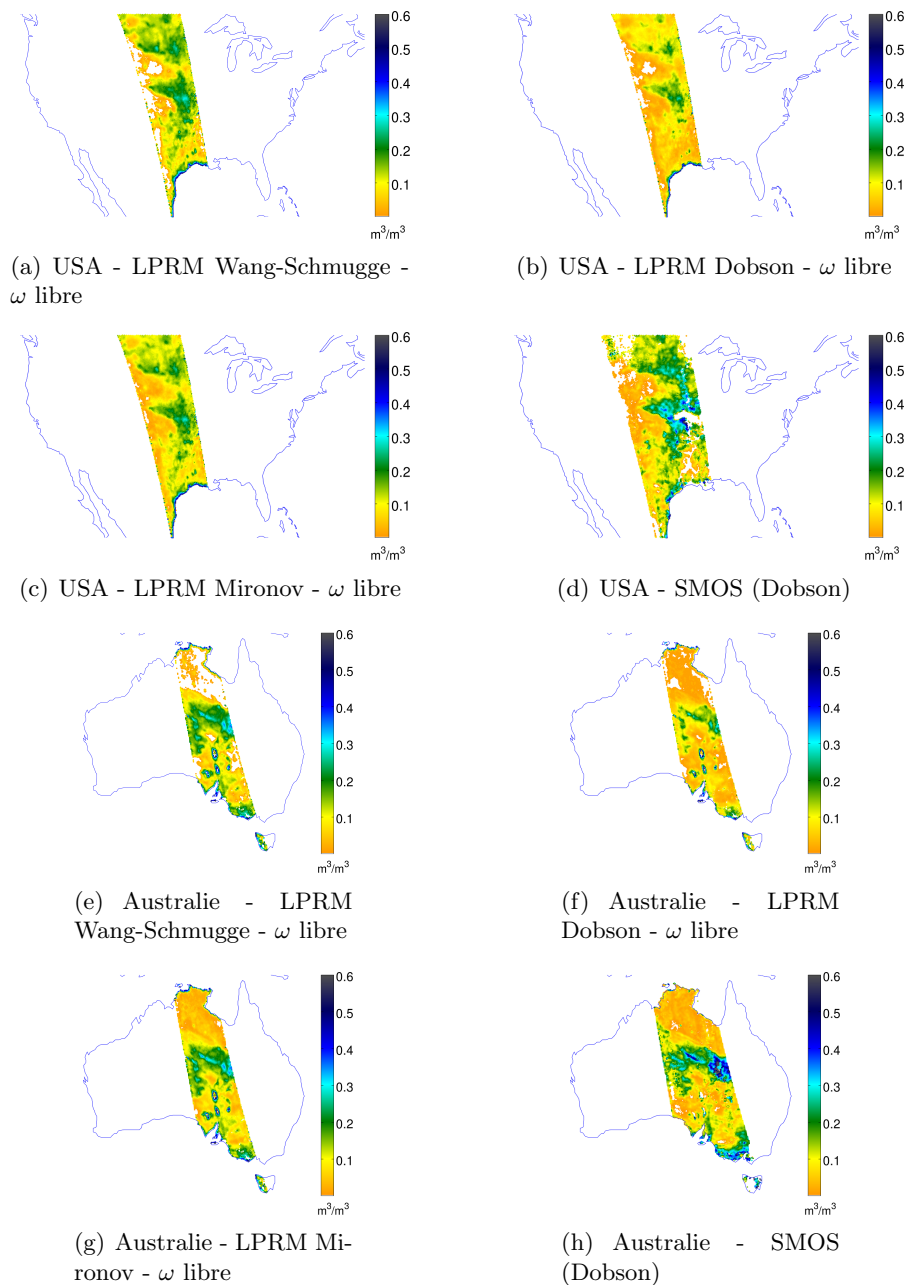


FIGURE 6.17 – Représentations spatiales des humidités retrouvées à partir des trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,e) LPRM avec le modèle de Wang-Schmugge, (b,f) LPRM avec le modèle de Dobson, (c,g) LPRM avec le modèle de Mironov et (d,h) SMOS avec le modèle de Dobson.

parties d'orbite aux États-Unis (16/09/2010) et en Australie (17/09/2010) alors que l'aspect temporel est étudié à travers les exemples de Little Washita et de Walnut Gulch en 2010.

En libérant la paramètre d'albédo ω , les humidités retrouvées sont plus faibles (fig. 6.17) que lorsque ω était fixé à 0 (fig. 6.13). Les humidités ainsi retrouvées sont plus proches de celles de SMOS. De plus on observe la même distribution spatiale de l'humidité.

Les trois modèles de constante diélectrique amènent à retrouver des humidités plus faibles que celles de SMOS (fig. 6.18, biais négatifs) mais le modèle de Dobson est celui qui fournit les valeurs les plus faibles (biais de $-0.07 \text{ m}^3/\text{m}^3$ sur les États-Unis et de $-0.04 \text{ m}^3/\text{m}^3$ sur l'Australie). En revanche, les modèles de Wang-Schmugge et de Mironov sont proches statistiquement. On peut remarquer que le modèle de Mironov permet de retrouver une valeur de SM sur un plus grand nombre de points surtout sur les faibles valeurs d'humidité (+15% sur les États-Unis et +30% sur l'Australie par rapport à Wang-Schmugge).

Les valeurs de ω optimisées en utilisant ces trois modèles de constante diélectrique sont proches (voir annexe F.1, fig. F.1). Sur les États-Unis, l'albédo est élevé avec des valeurs entre 0.1 et 0.3 et des pics supérieurs à 0.4. Sur l'Australie, les formes spatiales suivent celles de l'humidité avec de fortes valeurs au centre et plus faibles autour. Sur les deux bassins versants, les valeurs de l'albédo sont dispersées tout au long de l'année 2010. Sur Little Washita, ω est élevé en début d'année (autour de 0.2) puis baisse en Juillet-Août (autour de 0.1) et finit par osciller entre 0.1 et 0.2. Sur Walnut Gulch, on peut apercevoir une tendance ou une évolution plus nette de ce paramètre avec un niveau stable autour de 0.08 et deux augmentations en Février et Août jusqu'à 0.25. Dans l'algorithme qui permet de retrouver les humidités de SMOS, ω est fixé à 0 pour les zones dites nominales (végétation basse) et à 0.08 sur les forêts (ATBD, 2010). Les valeurs de ω retrouvées dans cette étude sont bien plus élevées que celles utilisées pour les humidités SMOS mais elles montrent un certain dynamisme spatial et temporel que SMOS ne prend pas en compte alors que cela impacte grandement les humidités retrouvées.

Ces trois modèles ont également été testés sur les séries temporelles de Little Washita (LW) et de Walnut Gulch (WG) en 2010 (fig. 6.19). Les humidités retrouvées sont bien inférieures à celles retrouvées avec $\omega=0$ (fig. 6.15) notamment lors des évènements pluvieux où il n'y a plus de surestimation automatique. Les modèles de Wang-Schmugge et de Mironov donnent des humidités très proches des mesures : biais de $-0.04 \text{ m}^3/\text{m}^3$ et $-0.05 \text{ m}^3/\text{m}^3$ sur LW et de $0.005 \text{ m}^3/\text{m}^3$ et $0.002 \text{ m}^3/\text{m}^3$ sur WG, fig. 6.20 ; alors que le modèle de Dobson sous-estime bien plus les humidités : biais de $-0.07 \text{ m}^3/\text{m}^3$ sur LW et de $-0.02 \text{ m}^3/\text{m}^3$ sur WG. Cependant, les trois modèles de constante diélectrique suivent la dynamique temporelle de l'humidité mesurée sur ces deux bassins versants.

Le modèle de Mironov permet encore une fois de retrouver une valeur de l'humidité sur un plus grand nombre de dates. Ce n'est pas clairement le cas sur LW mais sur WG, 50% de dates en plus sont retrouvées par rapport au modèle de Wang-Schmugge. Walnut Gulch est un site plus aride et plus sableux que Little Washita (Jackson et al., 2010) et il n'est pas surprenant de voir que Mironov se comporte mieux que les modèles de Wang-Schmugge ou de Dobson sous ces conditions.

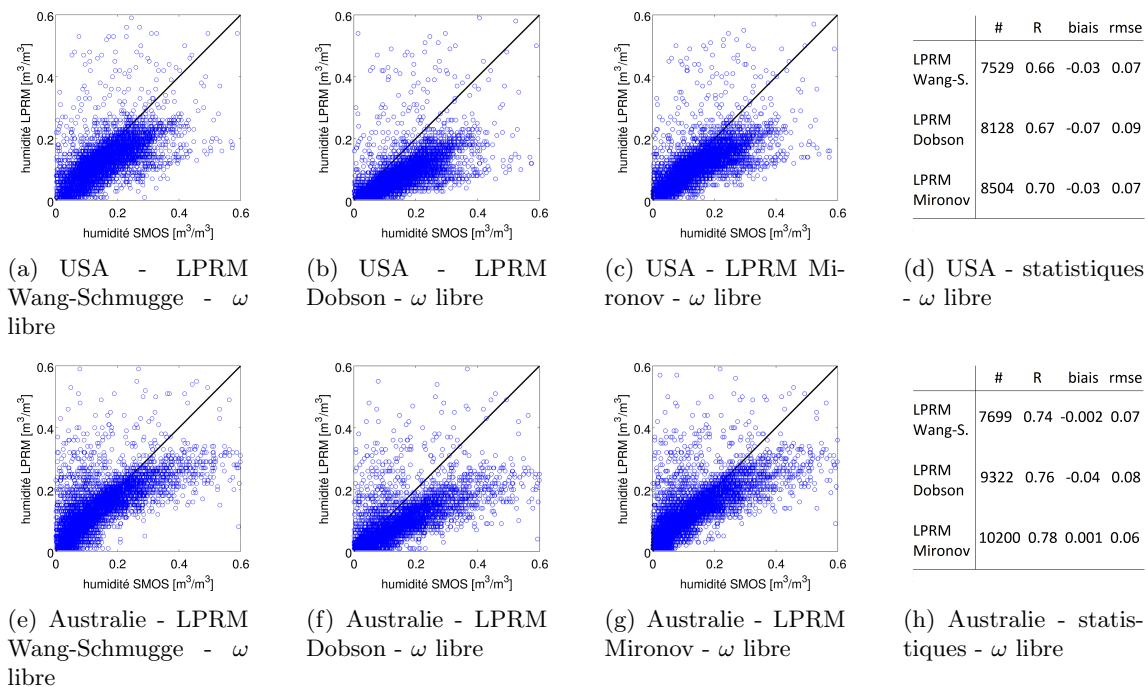


FIGURE 6.18 – Comparaison des humidités retrouvées avec le LPRM et les trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec celles de SMOS sur les Etats-Unis et sur l'Australie : (a,d) modèle de Wang-Schmugge, (b,e) modèle de Dobson et (c,f) modèle de Mironov.

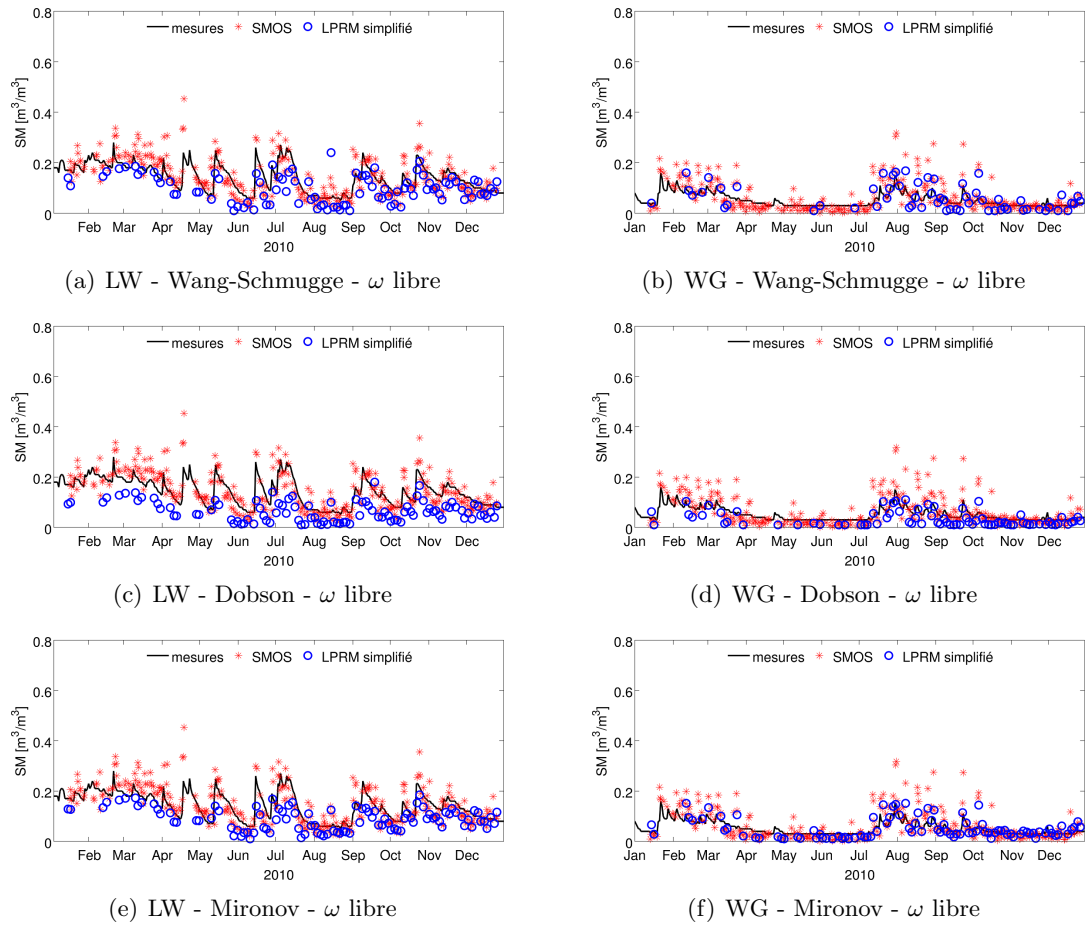


FIGURE 6.19 – Représentations temporelles des humidités retrouvées avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.

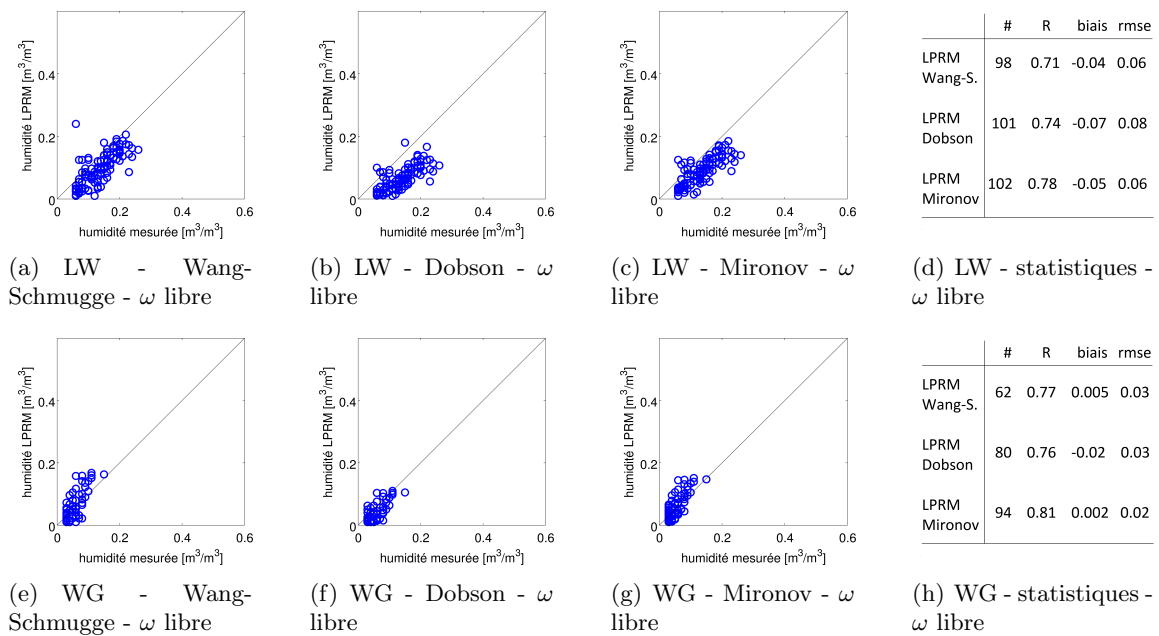


FIGURE 6.20 – Comparaison des humidités retrouvées par le modèle LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω sur Little Washita (LW) et Walnut Gulch (WG) : (a,e) Wang-Schmugge, (b,f) Dobson et (c,g) Mironov. Les statistiques sont présentées en (d,h).

6.3.3.2 Modèle G : rugosité optimisée et albédo nul

Si la rugosité du sol joue un rôle important à petite échelle, la rugosité est très souvent négligée lorsque l'on traite des informations satellite, d'où l'hypothèse $h=0$. Dans cette partie, l'albédo ω est fixé à 0 et le paramètre de rugosité est optimisé. Comme précédemment, deux paramètres du LPRM sont maintenant à optimiser : SM et h . La gamme de valeurs pour l'humidité reste la même ($0-0.6 \text{ m}^3/\text{m}^3$) et la rugosité est à optimiser entre 0 et 1.

Les résultats de cette étude (non montrés sous forme de figures) ont montré qu'en fixant ω à 0, la rugosité optimisée était toujours à 0 et que les résultats statistiques sont très proches de ceux du cas : $\omega=0$ et $h=0$. Ces résultats ont été les mêmes pour les États-Unis, l'Australie et les deux bassins versants Little Washita et Walnut Gulch.

Pour un albédo nul, la meilleure valeur pour la rugosité est donc 0. Cependant, plusieurs études montrent que la rugosité, même à l'échelle d'un champ de vue de satellite, n'est pas négligeable et varie avec l'humidité du sol SM (Escorihuela et al., 2007, Panciera et al., 2009).

6.3.3.3 Modèle H : albédo optimisé et dépendance de la rugosité à l'humidité de surface

Dans cette section, l'albédo ω redevient un paramètre à optimiser et la rugosité est fonction de l'humidité du sol. Plusieurs modélisations ont été développées (Escorihuela et al., 2007, Panciera et al., 2009, ATBD, 2010) et celle proposée par de Jeu et al. (2009) où h dépend de l'angle et de l'humidité est testée dans cette partie :

$$h = 0.4 - SM \cdot \theta^{1.5} \quad (6.18)$$

Cette modélisation a été déduite à partir du jeu de données NAFE'05 (Panciera et al., 2008) qui regroupe des mesures sur le terrain ainsi que des observations aéroportées sur un bassin versant au Sud-Est de l'Australie. Il faut noter que ces mesures ont été réalisées dans des conditions semi-arides et que cette modélisation n'est peut être pas adaptée pour des zones plus tempérées.

Comme dans les deux sections précédentes, deux paramètres sont à optimiser ici : l'humidité SM (entre 0 et $0.6 \text{ m}^3/\text{m}^3$) et l'albédo ω (entre 0 et 1). La rugosité dépendant maintenant de l'angle d'observation, celle montrée sur les figures en annexes F.2 est la valeur modélisée à 42.5° .

En introduisant une dépendance de la rugosité à l'humidité du sol, les humidités retrouvées (fig. 6.21) sont un peu plus élevées que lorsque la rugosité était négligée (fig. 6.17). Lorsque la rugosité augmente, les températures de brillance sont plus élevées (fig. 6.4 (a)) et si h est fixé, cette augmentation des températures se traduit par une diminution de l'humidité (fig. 6.2). Sur les États-Unis et l'Australie, le LPRM avec le modèle de Dobson est très proche de ce que produit SMOS avec un biais de $-0.02 \text{ m}^3/\text{m}^3$ sur les États-Unis et de $-0.005 \text{ m}^3/\text{m}^3$ sur l'Australie avec un nombre de points valides très proche de celui de Mironov (fig. 6.22). Les trois modèles sont en général très proches des humidités SMOS sur les faibles valeurs d'humidité ($0-0.2 \text{ m}^3/\text{m}^3$) puis donnent des valeurs plus faibles que celles de SMOS sur les événements pluvieux par exemple.

Ceci est confirmé par les humidités retrouvées sur les deux sites test aux États-Unis (fig. 6.23) où les humidités SMOS sont systématiquement plus élevées lors des événements pluvieux alors que le LPRM reste plus proche des mesures terrain. Sur LW, le modèle de Wang-Schmugge délivre des humidités plus proches des mesures avec un biais de $-0.003 \text{ m}^3/\text{m}^3$, alors que sur WG, le modèle de Mironov est le plus proche avec un biais de $0.001 \text{ m}^3/\text{m}^3$. Mironov est un modèle connu pour mieux fonctionner sur les sols sableux que modèles de Wang-Schmugge et de Dobson.

Les albédos retrouvés avec cette version du LPRM sont quasiment identiques à ceux retrouvés en négligeant la rugosité (voir annexes F.1 et F.2). Quant à la rugosité, elle évolue

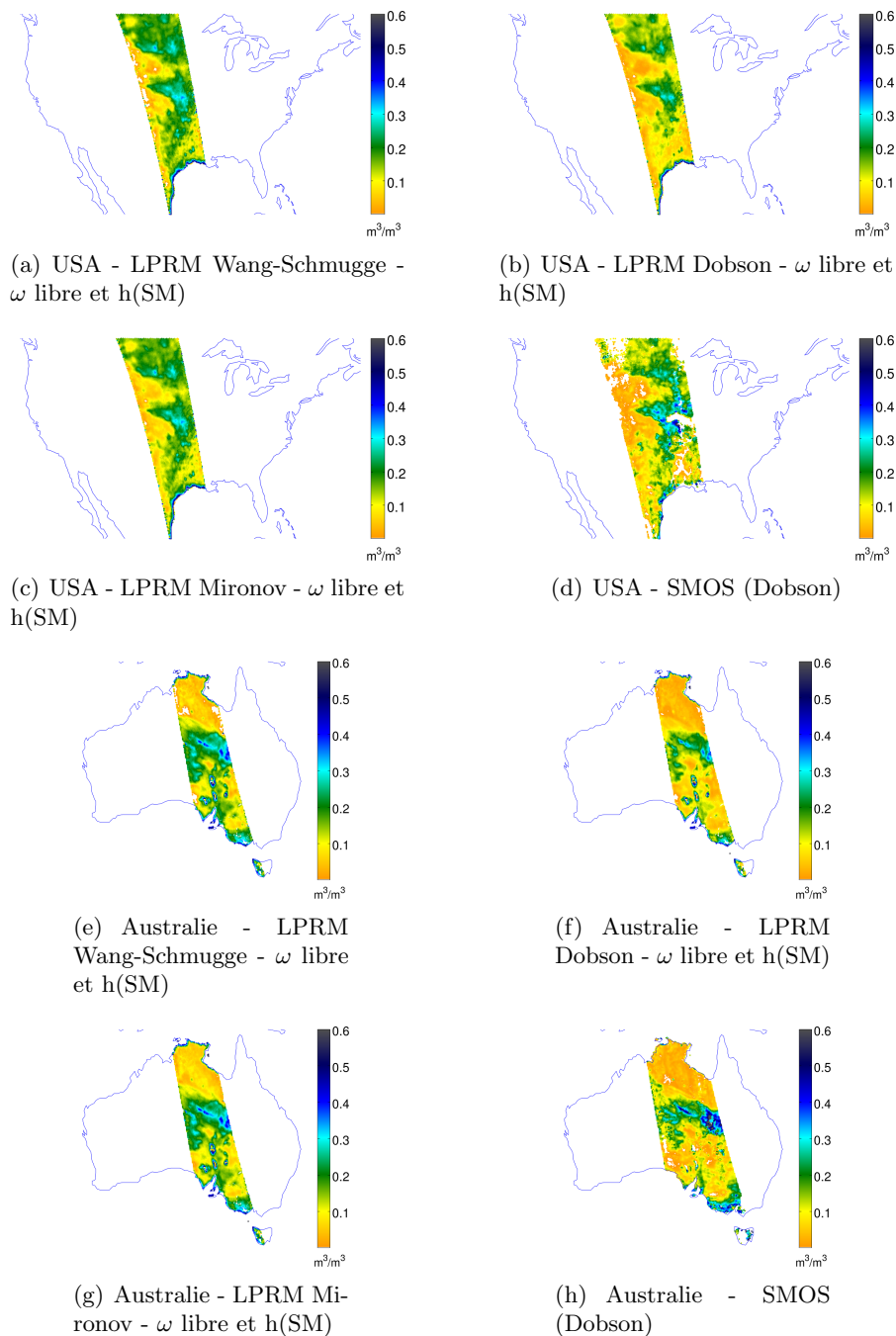


FIGURE 6.21 – Représentations spatiales des humidités retrouvées à partir des trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,e) LPRM avec le modèle de Wang-Schmugge, (b,f) LPRM avec le modèle de Dobson, (c,g) LPRM avec le modèle de Mironov et (d,h) SMOS avec le modèle de Dobson.

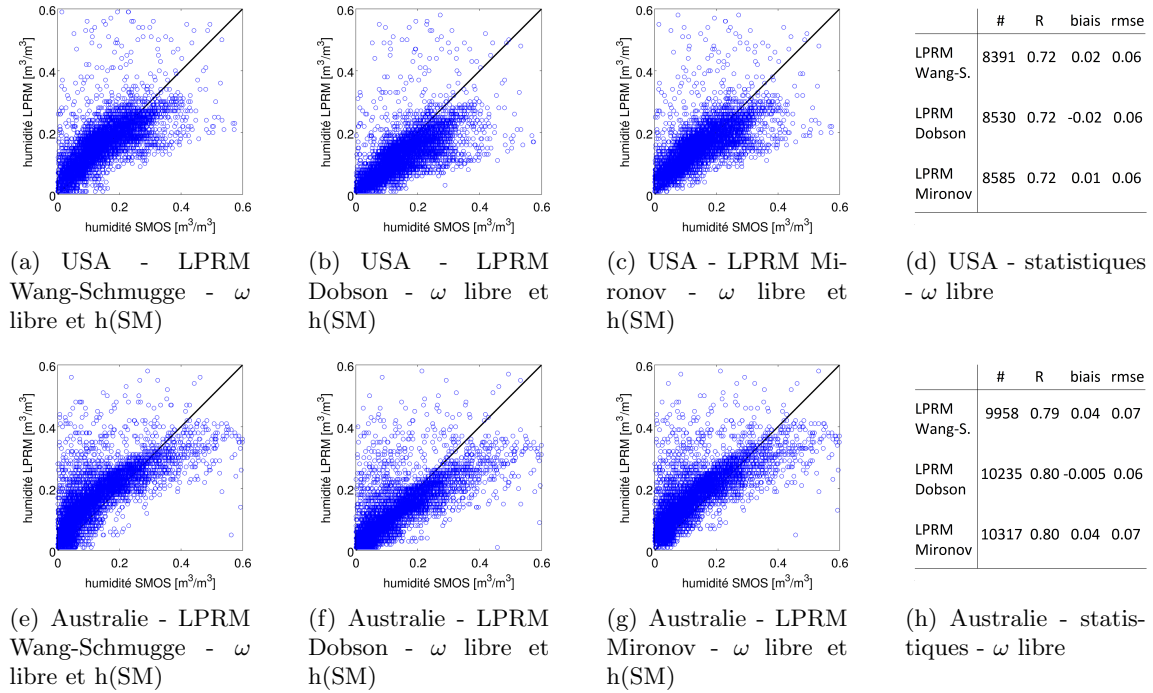


FIGURE 6.22 – Comparaison des humidités retrouvées avec le LPRM et les trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol avec celles de SMOS sur les Etats-Unis et sur l'Australie : (a,d) modèle de Wang-Schmugge, (b,e) modèle de Dobson et (c,f) modèle de Mironov.

entre 0.2 et 0.4 (0.4 étant la limite supérieure du modèle utilisé pour une humidité nulle). La répartition spatiale est fortement liée à celle de l'humidité à cause du modèle de rugosité utilisé.

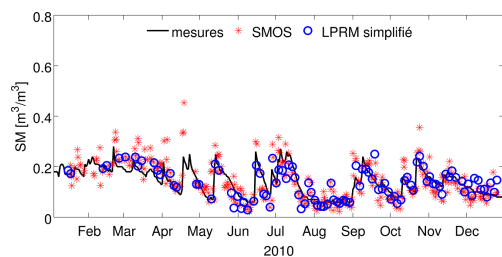
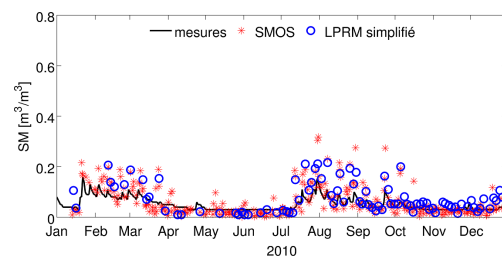
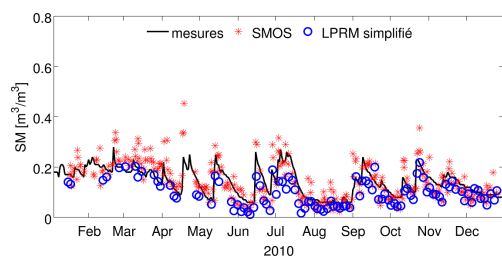
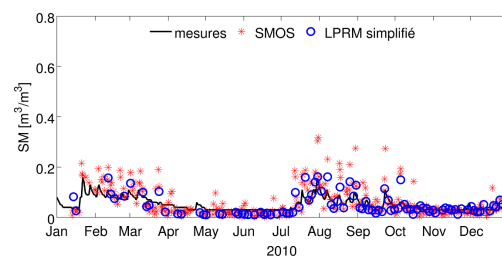
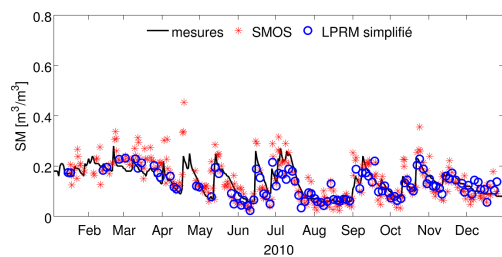
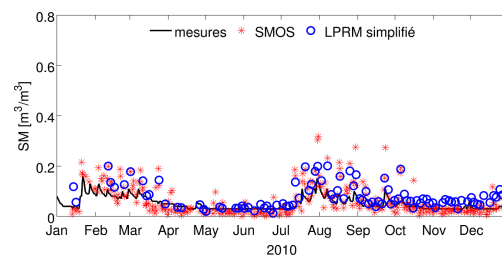
(a) LW - Wang-Schmugge - ω libre et h(SM)(b) WG - Wang-Schmugge - ω libre et h(SM)(c) LW - Dobson - ω libre et h(SM)(d) WG - Dobson - ω libre et h(SM)(e) LW - Mironov - ω libre et h(SM)(f) WG - Mironov - ω libre et h(SM)

FIGURE 6.23 – Représentations temporelles des humidités retrouvées avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.

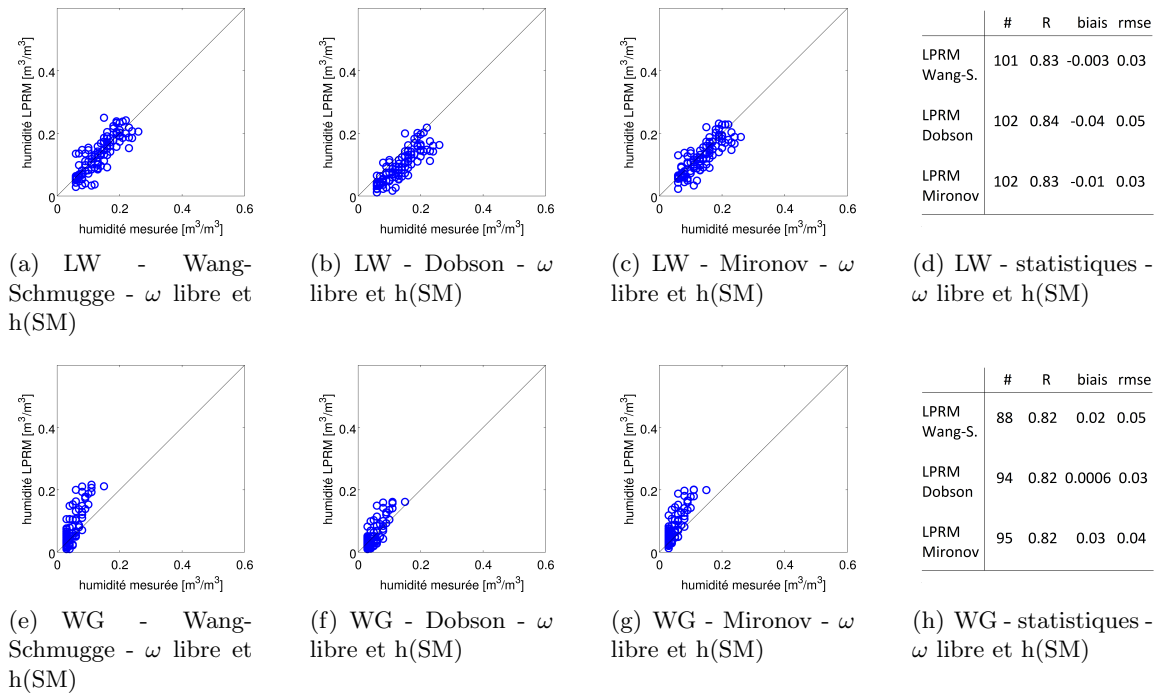


FIGURE 6.24 – Comparaison des humidités retrouvées par le modèle LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol sur Little Washita (LW) et Walnut Gulch (WG) : (a,e) Wang-Schmugge, (b,f) Dobson et (c,g) Mironov. Les statistiques sont présentées en (d,h).

6.3.4 Comparaison des épaisseurs optiques de la végétation

L'algorithme LPRM retrouve l'humidité du sol mais aussi l'épaisseur optique de la végétation. Ce paramètre n'est en revanche pas optimisé comme l'humidité mais est retrouvé en fonction du rapport de polarisation, de l'angle d'observation et des caractéristiques du sol dont l'humidité. Pour rappel, l'humidité et l'épaisseur optique de la végétation de SMOS sont retrouvées par optimisation.

Les épaisseurs optiques τ retrouvées par 5 options du modèle LPRM sont comparées par la suite sur les États-Unis et l'Australie :

- (a) LPRM avec Wang-Schmugge, $\omega = 0$ et $h=0$ (LPRM basique),
- (b) LPRM avec le rapport de polarisation (RP) lissé,
- (c) LPRM itératif,
- (d) LPRM avec Wang-Schmugge, ω libre et $h=0$, et
- (e) LPRM avec Wang-Schmugge, ω libre et h fonction de l'humidité SM.

Ces épaisseurs optiques retrouvées sont comparées à celles inversées par SMOS et à des cartes de végétation existantes (figs. 6.25 et 6.26).

Au fur et à mesure des tests effectués avec le LPRM, l'épaisseur optique a diminué en terme de valeur. Seul le LPRM utilisé avec le RP lissé ne permettait pas de retrouver une valeur sur tous les points disponibles. Le LPRM basique et itératif sont très proches. On peut d'ailleurs observer au Nord des États-Unis le fleuve Missouri traversant la fauchée d'Ouest en Est avec une épaisseur optique proche de 0.

Une première diminution du τ moyen est visible lorsque le paramètre d'albédo ω est libéré, surtout au Sud des États-Unis et de l'Australie. Mais c'est surtout lors de l'introduction de la dépendance de la rugosité à l'humidité que le τ moyen a vraiment diminué. Ceci peut s'expliquer en observant les profils de températures de brillance sur les figures 6.4 et 6.5 : les valeurs de SM et donc de τ étaient artificiellement modifiées pour pouvoir compenser les effets de l'albédo (qui était fixé à 0 puis libéré) et de la rugosité (qui était fixée à 0 puis libérée) sur les températures de brillance.

Aux États-Unis, on peut ainsi retrouver quelques structure de végétation avec les plaines au Nord et au centre avec une épaisseur optique plus faible alors qu'au Sud de la fauchée, il y a de plus en plus d'arbres et de forêts et cela se traduit par une augmentation de la valeur de τ .

En Australie, on peut également retrouver certaines structures avec une forêt dense au Nord et un peu au centre, entourée de désert alors qu'au Sud on retrouve cette forêt près des côtes.

Ces résultats et comparaisons ne sont qu'une première étape dans l'analyse de l'épaisseur optique de la végétation. Une des méthodes de validation serait de comparer ces résultats à l'indice NDVI.

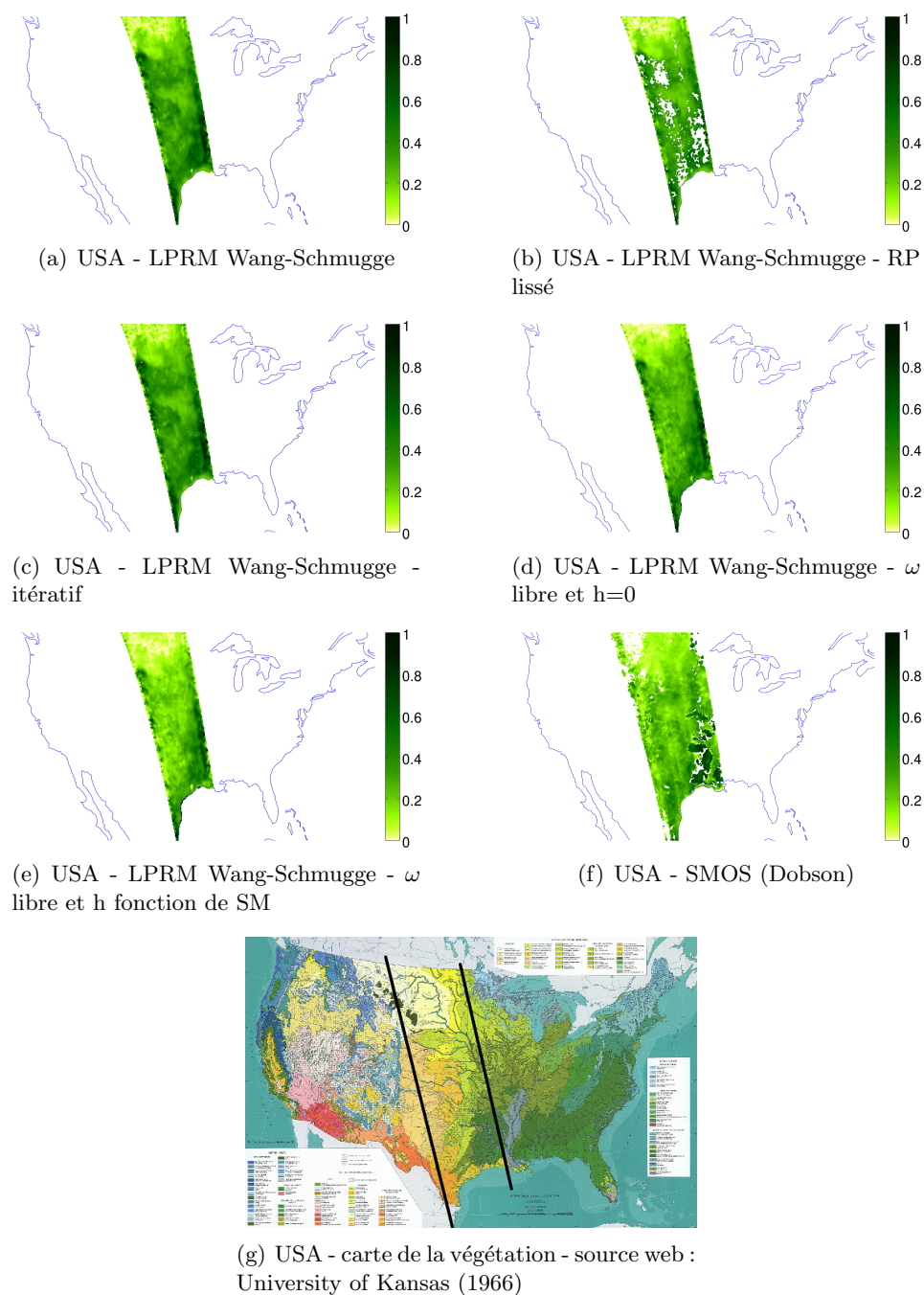


FIGURE 6.25 – Représentations spatiales des épaisseurs optiques de la végétation retrouvées.

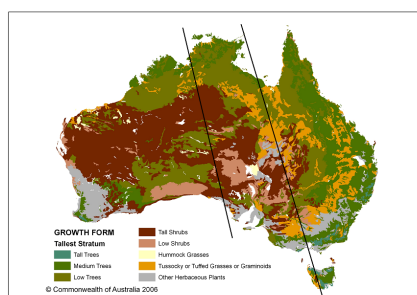
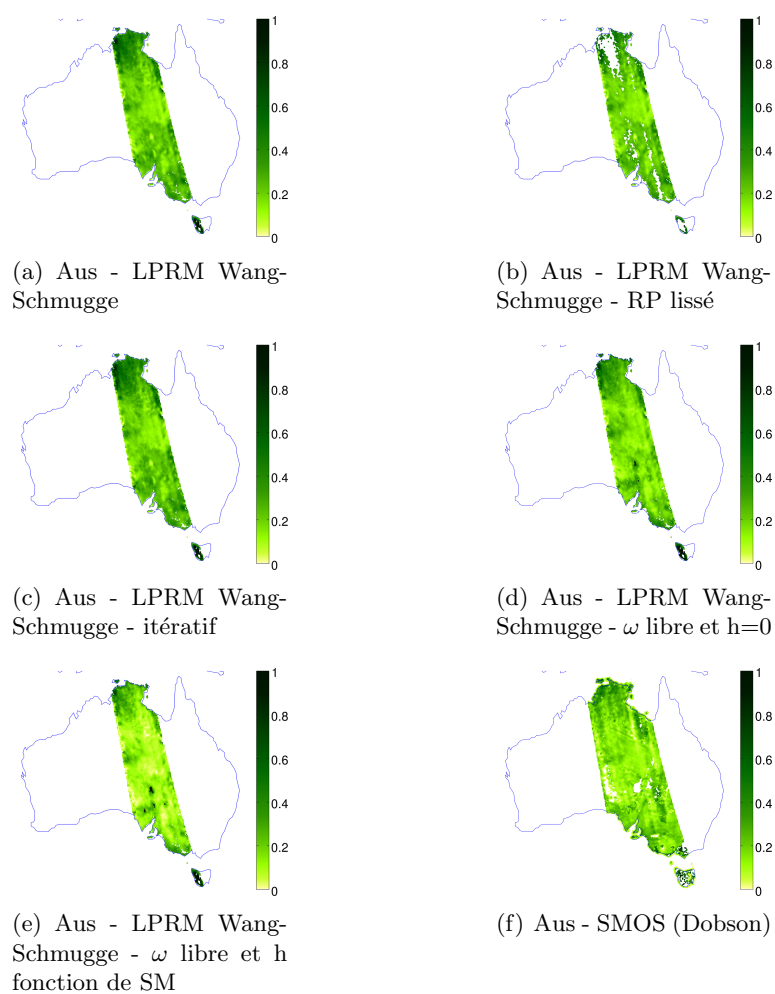


FIGURE 6.26 – Représentations spatiales des épaisseurs optiques de la végétation retrouvées.

6.4 Conclusions

Dans cette étude, le modèle simplifié LPRM (*Land Parameter Retrieval Model*, Owe et al. (2001)) pour retrouver l'humidité a été adapté pour l'utilisation des observations multi-angulaires en bande L de SMOS. Auparavant, ce modèle avait été utilisé avec les acquisitions de SMMR (*Scanning Multichannel Microwave Radiometer*), SSM/I (*Special Sensor Microwave/Imager*), TRMM (*Tropical Rainfall Measuring Mission*) ou encore AMSR-E (*Advanced Microwave Scanning Radiometer - Earth observation system*). Avec un algorithme unique, un historique de l'humidité à l'échelle globale avait été construit de 1978 à 2011 (arrêt d'AMSR-E, Owe et al. (2001)). Dans cet esprit de continuité, il était important de pouvoir adapter ce modèle aux nouvelles acquisitions réalisées par le radiomètre du satellite SMOS.

SMOS est la première mission à observer la Terre en bande L en vue de retrouver l'humidité des sols. La première étape a donc été d'adapter les valeurs des différents paramètres à la bande L. L'étude de de Jeu et al. (2009) sur le jeu de données NAFE(05 récoltées en zone semi-aride en Australie a été utilisée comme point de départ avec des propositions de valeurs pour l'albédo et la rugosité qui ont été utilisées dans ce travail.

La deuxième étape a été d'intégrer l'aspect multi-angulaire. En effet, le LPRM a été originellement conçu pour ne traiter que des données mono-angulaires. Cette multi-angularité a donc été introduite dans le processus d'optimisation qui est maintenant réalisé sur le profil angulaire entier et non plus sur un seul angle. La valeur d'humidité retrouvée est celle qui réduit au maximum la somme des différences entre les températures de brillance (TB) modélisées et observées.

L'analyse des profils angulaires des TB modélisées a révélé une limite concernant la manière de calculer l'épaisseur optique τ de la végétation. Dans le LPRM original, τ est calculé pour chaque angle à partir du rapport de polarisation (RP) et de la valeur d'humidité SM. Cela se traduit par un profil angulaire des TB modélisées discontinu. Deux approches ont alors été testées pour rétablir la continuité des TB : lisser le RP par rapport à l'angle d'observation et retrouver le couple de valeurs (SM, τ) itérativement. Ces trois approches donnent des résultats équivalents sur des profils angulaires de TB non bruités mais s'avèrent très différentes sur des profils plus bruités avec des valeurs d'humidité retrouvées pouvant doubler. Le modèle original a été gardé pour la suite de l'étude.

Le second point d'intérêt de ce travail porte sur le choix du modèle de constante diélectrique. Trois modèles ont été testés : Wang & Schmugge (1980), Dobson et al. (1985) et Mironov et al. (2004). Le modèle de Wang-Schmugge est implémenté dans le LPRM original alors que le modèle de Dobson est utilisé dans le niveau 2 de SMOS (v400) qui permet de retrouver l'humidité et le modèle de Mironov est implémenté dans le nouveau prototype de SMOS. Les trois modèles surestiment l'humidité par rapport à SMOS et aux mesures terrain, notamment lors des événements pluvieux. Le modèle de Dobson est cependant le modèle qui a permis de se rapprocher le plus des mesures terrain.

La dernière partie de cette étude s'est penchée sur la libération des paramètres de rugosité h et d'albédo ω . Originellement, ces paramètres sont fixes dans le LPRM. La libération de la rugosité n'a pas apporté d'amélioration puisque h convergeait quasiment partout et tout le temps vers 0. La libération et l'optimisation du paramètre ω ont permis de retrouver des valeurs d'humidité plus basses qui se rapprochent des humidités SMOS et des mesures terrain, voire qui les sous-estiment. L'introduction de la dépendance de la rugosité à l'humidité a permis de retrouver les valeurs d'humidité les plus proches des mesures terrain confirmant une nouvelle fois cette dépendance (Escorihuela et al., 2007, Saleh et al., 2009).

Cette étude sur un algorithme plus simple que celui utilisé pour le niveau 2 de SMOS permet de mettre en avant plusieurs axes de recherche futurs :

- intérêt d'avoir le choix entre plusieurs modèles de constante diélectrique : Wang-Schmugge et Dobson étaient parfois très proches mais Mironov peut apporter une nouvelle modélisation plus avantageuse notamment sur les sols très secs et sableux où plus de valeurs d'humidité valides sont retrouvées,
- libération de l'albédo ω : dans cette étude, la répartition spatiale de ω lors de son optimisation n'était pas aléatoire et suivant certaines formes spatiales et un des futurs axes d'amélioration du niveau 2 de SMOS serait de définir une valeur de ω pour chaque type de végétation,
- l'introduction de la dépendance de la rugosité h à l'humidité SM a réellement permis d'améliorer l'algorithme simplifié. SMOS le fait déjà et son modèle peut encore être amélioré.

Cet algorithme simplifié reste encore à améliorer mais pourrait être implémenté dans le niveau 2 de SMOS comme phase d'initialisation, aussi bien pour l'humidité SM que pour l'épaisseur optique de la végétation τ . La deuxième possibilité serait de l'utiliser comme deuxième algorithme si l'algorithme SMOS ne convergeait pas, et en ne tenant donc plus compte de l'hétérogénéité de la surface observée.

Une autre étude possible serait de quantifier l'apport du multi-angulaire. Il s'agirait de ne garder qu'un seul angle (à déterminer) et d'utiliser cet algorithme afin de retrouver l'humidité du sol et l'épaisseur optique de la végétation si possible, et de comparer ces inversions avec celles des observations multi-angulaires. Ainsi on pourrait déterminer les avantages et les inconvénients (si le profil angulaire est très bruité par exemple) d'optimiser sur une gamme entière d'angles d'observation.

Conclusions générales

L'eau est un élément essentiel de la vie sur Terre et avec le réchauffement climatique qui ne cesse de s'accélérer, son futur représente une grande source d'interrogation et d'inquiétude. Les changements climatiques ont commencé à se manifester par une augmentation de la température du sol, une élévation du niveau moyen des océans ou encore par une désertification de certaines régions (IPCC - *Intergovernmental Panel on Climate Change*, 2007). L'impact de ce dérèglement sur le cycle de l'eau reste encore inconnu. Afin d'observer ces changements climatiques, le GCOS (*Global Climate Observing System*) a sélectionné 50 variables climatiques essentielles (World Meteorological Organization et al., 2010). L'humidité du sol, ou contenu en eau du sol, fait partie des variables à surveiller pour suivre l'évolution du cycle hydrologique sur les terres émergées. L'humidité du sol ne décrit qu'une infime partie du cycle de l'eau. Cette variable reste cependant très importante puisqu'elle représente l'eau disponible pour la croissance de la végétation et des cultures. L'enjeu le plus important est donc de bien quantifier et gérer cette réserve d'eau.

De nombreux satellites ont déjà permis de retrouver l'humidité des sols à l'échelle globale dont AMSR-E (*Advanced Microwave Scanning Radiometer-Earth observation system*, 2002-2011) ou ASCAT (*Advanced Scatterometer*, 2006-). De toutes les missions spatiales déjà lancées, aucune n'avait été spécialement conçue pour retrouver l'humidité des sols et leurs instruments opéraient tous à des fréquences supérieures à 6 GHz. La mission SMOS (*Soil Moisture and Ocean Salinity*, Kerr et al. (2001)) est donc la première mission spatiale qui permet de retrouver l'humidité en observant la Terre à la fréquence idéale pour cette variable : 1.4 GHz (Kerr et al., 2001, 2010). Même si les satellites précédents n'utilisaient pas cette fréquence optimale, cette série temporelle d'informations reste d'une grande valeur.

SMOS embarque à son bord un radiomètre interféromètre composé de 69 antennes disposées sur trois bras. Ce dispositif permet d'observer pour la première fois la Terre en bande L sous de multiples angles (de 0° à 55°). Au delà de la nouveauté conceptuelle, SMOS apporte un algorithme novateur pour retrouver l'humidité et l'épaisseur optique de la végétation à travers la prise en compte de l'hétérogénéité de la scène observée. Un des objectifs de cette mission est d'estimer l'humidité avec une précision de 0.04 m³/m³ en couvrant le globe tous les trois jours.

L'historique de l'humidité des sols des trente dernières années est donc disponible à travers les observations des satellites micro-ondes et la mission SMOS vient s'ajouter à cette série de satellites. Dans un contexte de changement climatique, il est primordial d'avoir accès à de longues séries temporelles des variables physiques d'intérêt afin d'étudier leur évolution à travers le temps. Cet historique de l'humidité n'est cependant pas utilisable pour des études climatiques

car elle présente de nombreuses discontinuités lors des changements de capteurs notamment où des sauts voire des dynamiques opposées apparaissent. Il est donc indispensable d'harmoniser ces séries temporelles d'humidité afin qu'elles puissent être utilisables ensemble.

Les objectifs de cette thèse sont multiples : (1) validation des données SMOS aussi bien à l'échelle locale qu'à l'échelle globale, (2) étude statistique globale permettant d'identifier les sources des erreurs des produits d'humidité SMOS, (3) proposition d'une nouvelle méthode statistique permettant l'homogénéisation temporelle de plusieurs produits d'humidité, (4) création d'un modèle simplifié pour retrouver l'humidité à partir des acquisitions SMOS.

Résultats

La première partie de cette thèse a été consacrée à la validation du produit d'humidité SMOS à l'échelle locale dans un premier temps puis à l'échelle globale. Quatre bassins versants ont été sélectionnés aux États-Unis pour la première phase de validation : Walnut Gulch, Little Washita, Little River et Reynolds Creek. Ces quatre sites test sont tous équipés d'un réseau de stations qui mesurent l'humidité à 5 cm de profondeur. Ces réseaux ont été conçus initialement pour la validation des observations AMSR-E. Ils sont donc adaptés à la taille d'un champ de vue de satellite.

Une comparaison de plusieurs produits d'humidité a été réalisée sur ces bassins versants : SMOS, AMSR-E avec l'algorithme de VUA (*Vrije Universiteit Amsterdam*, Owe et al. (2001)), ASCAT et le modèle ECMWF (centre européen de prévisions météorologiques à moyen terme). Aucun produit n'a obtenu une corrélation significativement meilleure qu'un autre. En revanche, l'humidité retrouvée par le modèle ECMWF est la moins dispersée mais présente un biais non négligeable et constant. L'indice d'humidité d'ASCAT est très dispersé et ne représente quasiment aucune dynamique comparé aux mesures terrain. Le produit d'humidité d'AMSR-E/VUA a une bonne dynamique mais présente un biais variable selon les saisons que l'on suppose être relié à la végétation. L'humidité SMOS est le meilleur produit en terme d'erreur avec $0.052 \text{ m}^3/\text{m}^3$ en moyenne pour les orbites du matin et $0.073 \text{ m}^3/\text{m}^3$ pour l'après-midi. Même si SMOS n'atteint pas l'objectif initial de précision ($0.04 \text{ m}^3/\text{m}^3$), ce nouveau produit améliore d'un facteur 2 minimum l'erreur par rapport aux autres produits d'humidité testés avec une corrélation équivalente.

La validation locale est une première étape dans une phase d'évaluation puisque les sites testés sont bien trop spécifiques à des régions et des climats particuliers pour pouvoir dessiner une conclusion générale. La deuxième étape est la validation à l'échelle globale. La méthode statistique de triple collocation a été appliquée aux produits d'humidité suivants sur 2010 : AMSR-E (VUA), AMSR-E (NSIDC), ASCAT, ECMWF et SMOS. Cette méthode ne permet pas d'obtenir des erreurs absolues mais des cartes d'erreurs relatives au choix des trois jeux de données auxquels la méthode est appliquée. Les humidités SMOS ont été comparées tour à tour à deux autres produits parmi ceux cités plus haut.

Il a été possible d'identifier les régions du monde où SMOS avait les plus faibles erreurs et celles où les erreurs étaient plus élevées. L'Amérique du Nord, l'Australie et l'ouest de l'Asie sont les régions où SMOS obtient de meilleurs résultats que les autres produits d'humidité. Afin d'analyser et de comprendre ces cartes d'erreur, les méthodes statistiques ANOVA (analyse de la variance) et CART (classification et arbre de régression) ont été utilisées afin d'identifier les sources de ces erreurs parmi les paramètres suivants : probabilité d'occurrence des RFI (interférences) en 2010, texture du sol (pourcentages de sable et d'argile), fraction de forêt observée par le radiomètre, fraction de zone humide, fraction d'eau libre, fraction d'eau salée,

fraction de sol nu, fraction de glace et fraction de zone urbaine.

A l'échelle globale, la fraction de forêt vue par le radiomètre explique la majeure partie de la variance de l'erreur SMOS (57%) suivi de la texture du sol (22%). Plus le radiomètre voit de la forêt, plus l'erreur est grande. Ces proportions varient grandement selon les continents. Lorsque la fraction de forêt explique 67% de l'erreur en Amérique du Nord, les interférences expliquent 47% de l'erreur en Asie centrale. La classification des erreurs a également mis en avant le rôle de la forêt avec des valeurs seuil très faibles (maximum de 7% de forêt visible dans le champ de vue de l'instrument). Même une très petite fraction de forêt observée par le radiomètre peut donc mener à une valeur très élevée de l'erreur SMOS.

Le produit d'humidité SMOS a ensuite été replacé dans un contexte plus temporel. De nombreux satellites ont déjà permis de retrouver l'humidité des sols à l'échelle globale depuis 1978 et SMOS fait maintenant partie de cette série de satellites. L'historique de l'humidité des sols des trente dernières années créé par l'Université d'Amsterdam (Owe et al., 2001) comporte de nombreuses discontinuités, notamment lors des changements de capteurs. Deux méthodes statistiques sont proposées pour homogénéiser des séries temporelles : le CDF matching (*Cumulative Density Function*) et les copules. Ces deux méthodes ont été appliquées aux humidités retrouvées par AMSR-E (VUA) et SMOS afin de créer une série temporelle continue de 2003 à 2010 en utilisant SMOS comme référence. L'année 2010 a été utilisée comme année de référence afin de calibrer les deux modèles statistiques.

Une série temporelle a ainsi été générée pour chaque site test de 2003 à 2010. Par rapport aux mesures terrain, le CDF matching et les copules gardent le même niveau de corrélation que les humidités originales d'AMSR-E mais diminuent l'erreur. Le CDF matching donne de bons résultats mais ne prend pas en compte la structure de dépendance entre les produits d'origine alors que les copules permettent de modéliser cette structure. Les copules fournissent donc de meilleurs résultats que le CDF matching aux valeurs extrêmes, comme après une pluie par exemple. La méthode des copules apporte une diminution de 25% en moyenne en terme d'erreur par rapport au CDF matching sur la longue série temporelle de 2003 à 2010. Les copules ont également l'avantage d'estimer l'incertitude des humidités retrouvées, ce qui est très apprécié dans des modèles d'assimilation de données. Les limitations sont les mêmes pour les deux méthodes : seule la dynamique de l'année de référence est prise en compte. Si un événement exceptionnel se produit une autre année alors il sera mal retranscrit par ces méthodes.

La dernière partie de cette thèse a été consacrée à l'adaptation de l'algorithme VUA (qui a déjà permis de générer le grand historique de l'humidité, Owe et al. (2001)) aux observations multi-angulaires en bande L de SMOS. Toujours dans cet esprit de continuité, il est important d'incorporer SMOS dans cette série en utilisant le même algorithme pour retrouver l'humidité des sols. Les humidités retrouvées par VUA sont sensiblement différentes de celles de SMOS puisque l'hétérogénéité de la surface observée n'est pas prise en compte par l'algorithme de VUA.

La première étape a été d'adapter les valeurs de certains paramètres du modèle à la bande L (de Jeu et al., 2009). La seconde étape a été la prise en compte de la multi-angularité des observations SMOS. Jusqu'à maintenant, les informations traitées par l'algorithme étaient mono-angulaires. Une optimisation sur l'intégralité du profil angulaire a ainsi été mise en place. Les humidités retrouvées présentaient des formes spatiales et des dynamiques temporelles très proches de celles de SMOS et des mesures terrain mais avaient également un fort biais (plus de $0.1 \text{ m}^3/\text{m}^3$).

De nouvelles approches ont été proposées concernant la modélisation de la constante diélectriques, l'albédo et la rugosité. Originellement, le modèle de VUA utilise la modèle de

constante diélectrique de Wang & Schmugge (1980). Les modèles de Dobson et al. (1985) (utilisé jusque maintenant par SMOS) et de Mironov et al. (2004) (très prochainement utilisé par SMOS) ont été implémentés et testés. Le modèle de Mironov permet de retrouver une valeur d'humidité sur plus de points, notamment dans les zones sèches et sableuses où le contenu en eau du sol est très faible. Le paramètre d'albédo était initialement fixé à 0 puis a été optimisé dans un second temps. Cela a entraîné une diminution globale des valeurs d'humidité retrouvées, diminuant ainsi le biais constaté auparavant. Le paramètre de rugosité était fixé à 0 et une dépendance à l'humidité et à l'angle d'observation a été introduite. Cette nouvelle paramétrisation a également permis de réduire le biais jusqu'à obtenir des erreurs très faibles du même ordre que celles de SMOS comparé aux mesures terrain.

Perspectives

La phase de validation des humidités SMOS a montré le potentiel de ce nouveau produit et a également permis de mettre en avant plusieurs axes de recherche.

Les humidités retrouvées par SMOS sont très proches des mesures terrain avec des biais très faibles et pour se rapprocher de l'objectif de précision de $0.04 \text{ m}^3/\text{m}^3$, il est nécessaire de diminuer la dispersion des valeurs d'humidité. L'un des axes d'amélioration serait une meilleure modélisation de la couche de végétation qui générerait une épaisseur optique moins bruitée. L'amélioration du modèle de la forêt est un point capital pour l'algorithme de niveau 2 qui prend en compte l'hétérogénéité des surfaces observées. Ce point fait partie des priorités et est l'objet de recherches plus approfondies par les équipes italiennes de Paolo Ferrazzoli à l'université Tor Vergata de Rome. Il est envisagé dans un futur proche que des cartes de textures plus précises soient utilisées dans le niveau 2 de SMOS. Ces cartes de textures seront régionales et fournies par les agences locales.

Un point qui n'a pas été couvert par la validation globale est l'influence de l'uniformité de la surface observée sur l'humidité retrouvée. Une scène contenant 20% de forêt et 80% de végétation basse peut être perçue différemment si la forêt est regroupée ou dispersée dans le champ de vue de l'instrument. Pour mener à bien cette étude, il faudrait d'abord définir un indice d'homogénéité avec plusieurs valeurs seuils et cette tâche, qui doit être réalisée en amont, n'est pas simple.

Les méthodes statistiques proposées pour homogénéiser l'historique de l'humidité des sols prenaient en compte la saisonnalité de cette variable en découpant l'année en quatre saisons. Les méthodes avaient ensuite été appliquées indépendamment sur chacune des saisons, ce qui pouvait créer des discontinuités au changement de saison. Utiliser une fenêtre glissante de 90 jours permettrait de s'affranchir de ces problèmes de transition. Une autre solution serait d'introduire le temps dans les formules des copules.

Les deux méthodes proposées ont été testées et validées sur quatre sites aux États-Unis et la prochaine étape serait de les utiliser à l'échelle globale. Le CDF matching peut rapidement être implémenté mais la méthode des copules demanderait une phase de travail en amont supplémentaire. En effet, le choix de la famille de copules, qui caractérise le lien entre les séries temporelles, doit être réalisé pour chaque point, ce qui rend cette méthode quasi-empirique. Comme déjà amorcé à la fin du chapitre 5, il faudrait pouvoir relier le choix de la famille de copules à certaines caractéristiques du sol (texture ou occupation).

De plus l'objectif à long terme de cette étude est de créer un historique homogène de l'humidité du sol de 1978 à nos jours en s'appuyant sur les observations SMOS. Dans cette étude, seule la relation entre AMSR-E et SMOS a été étudiée. La prochaine étape serait donc d'appliquer ces méthodes statistiques aux précédents satellites afin de compléter la série temporelle.

L'adaptation de l'algorithme de VUA reste encore à améliorer, notamment la méthode de calcul de l'épaisseur optique de la végétation avec des observations multi-angulaires, ou encore la relation entre la rugosité et l'humidité de surface. Cependant cet algorithme simple pourrait être implémenté dans le niveau 2 de SMOS comme phase d'initialisation, aussi bien pour l'humidité que pour la végétation. Il pourrait également être utilisé comme deuxième algorithme si celui de SMOS ne convergeait pas. Cette partie soulève également des interrogations concernant l'albédo dont une valeur constante n'est peut être pas adaptée. De plus amples études seraient nécessaires pour fixer des valeurs d'albédo différentes suivant les types de sols.

L'objectif de cette partie était également d'incorporer les observations de SMOS dans la longue série temporelle de l'humidité générée par l'algorithme de VUA. Dans le futur, il faudrait étudier la relation entre ce *nouveau* produit d'humidité et ceux d'AMSR-E. Le CDF matching et la méthode des copules permettraient ainsi de réaliser un historique de l'humidité entièrement homogène avec l'utilisation d'un algorithme unique.

De nouvelles missions spatiales sont en cours ou vont bientôt démarrer pour continuer l'historique de l'humidité des sols. On peut citer la mission Aquarius lancée en Juin 2011 ainsi que SMAP (*Soil Moisture Active Passive*) dont le lancement est prévu en Octobre 2014. Toutes les études menées et les méthodes développées lors de ce doctorat pourront être réutilisées pour la validation et la mise en commun de ces futurs produits d'humidité.



Intervalle de confiance

A.1 Intervalle de confiance pour une population à distribution normale

L'intervalle de confiance est le plus facilement déterminé pour une population à distribution normale (en forme de cloche) dont on cherche à déterminer la moyenne \bar{x} . Si N est la taille de l'échantillon X et que $\sigma(X)$ est son écart-type alors on estime à 68% la probabilité que la moyenne se situe dans l'intervalle suivant :

$$\left[\bar{x} - \frac{\sigma(X)}{\sqrt{N}}; \bar{x} + \frac{\sigma(X)}{\sqrt{N}} \right] \quad (\text{A.1})$$

et on l'estime à 95% pour l'intervalle :

$$\left[\bar{x} - 2 \frac{\sigma(X)}{\sqrt{N}}; \bar{x} + 2 \frac{\sigma(X)}{\sqrt{N}} \right] \quad (\text{A.2})$$

A.2 Intervalle de confiance pour un coefficient de corrélation

La distribution du coefficient de corrélation n'étant pas normale, il faut trouver un moyen de transformer cette variable afin de la normaliser. Pour cela on utilise la transformation dite de Fisher :

$$z = \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \quad (\text{A.3})$$

La nouvelle variable z ainsi créée est approximativement normale avec une moyenne de $1/2 \ln(1+R/1-R)$ et un écart-type de $1/\sqrt{N-3}$ (si la distribution originale de X est normale). La transformation inverse est donnée par :

$$R = \frac{\exp(2z) - 1}{\exp(2z) + 1} \quad (\text{A.4})$$

L'intervalle de confiance pour un coefficient de corrélation est calculé suivant la procédure suivante :

1. Calcul du coefficient de corrélation R de l'échantillon,
2. Transformation de R en z avec l'équation (A.3),

3. Utilisation d'une table pour trouver la valeur de z' pour laquelle le niveau de confiance est respecté (<http://stattrek.com/online-calculator/normal.aspx>). Pour un intervalle de confiance à 95%, on a $z' = 1.960$,
4. Calcul de l'écart-type $\sigma(z) = 1/\sqrt{N-3}$,
5. Limite inférieure $= z - z' \times \sigma(z)$,
6. Limite supérieure $= z + z' \times \sigma(z)$,
7. Passage des limites inférieure et supérieure de z vers R avec l'équation (A.4).

A.3 Intervalle de confiance pour un écart-type

Dans un premier temps, il faut définir l'estimateur sans biais de la variance (carré de l'écart-type) :

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (\text{A.5})$$

Ceci est la définition de l'estimateur *sans biais* de la variance. Il existe un autre estimateur où $N-1$ est remplacé par N mais il comporte un biais. L'utilisation de $N-1$ au lieu de N est connue sous le nom de correction de Bessel. De plus, $(N-1)$ correspond au nombre de degré de liberté du vecteur des résidus $(x_1 - \bar{x}, \dots, x_N - \bar{x})$.

Le théorème de Cochran montre que la variable $(N-1)s^2/\sigma^2$ suit une loi de χ^2 à $N-1$ degrés de liberté. Ce résultat est fondamental puisqu'il va permettre de calculer les bornes de l'intervalle de confiance de l'écart-type avec :

$$\sqrt{\frac{(N-1)s^2}{\chi_{\text{right}}^2}} \leq \sigma \leq \sqrt{\frac{(N-1)s^2}{\chi_{\text{left}}^2}} \quad (\text{A.6})$$

Les valeurs χ_{right}^2 et χ_{left}^2 sont trouvées à partir de tables prédéfinies à 2 entrées (comme par exemple <http://people.richland.edu/james/lecture/m170/tbl-chi.html> ou en plus interactif <http://alain.pichereau.pagesperso-orange.fr/chi2test.html>, section 4-2) :

1. le niveau de confiance : pour avoir un intervalle de confiance à 95%, le niveau de confiance α est de 0.05,
2. la taille de l'échantillon N .

Ainsi on définit :

$$\chi_{\text{right}}^2 = \chi^2(\alpha/2, N-1) \quad (\text{A.7})$$

$$\chi_{\text{left}}^2 = \chi^2(1 - \alpha/2, N-1) \quad (\text{A.8})$$

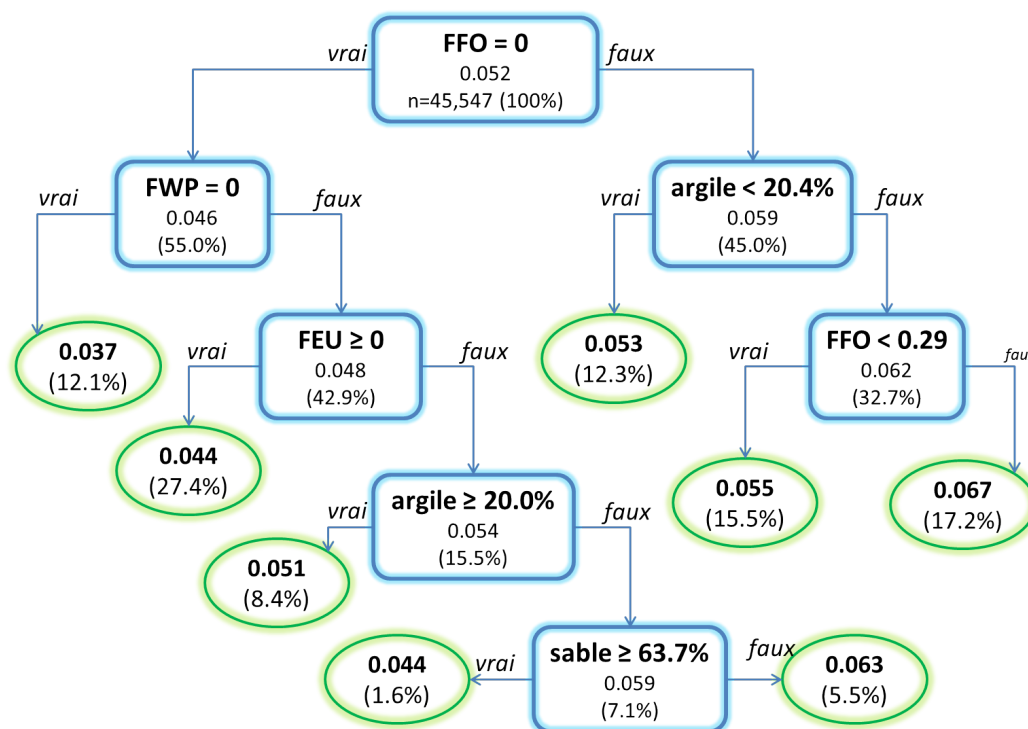
Annexe **B**

Arbres de régression pour chaque continent

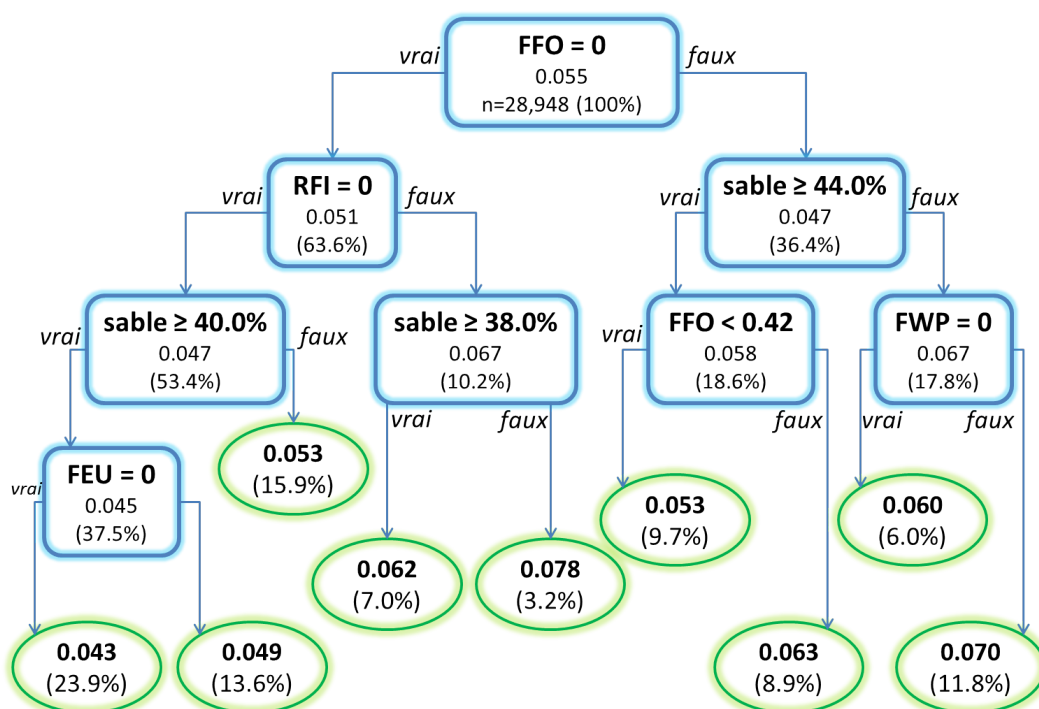
Des cartes des erreurs SMOS de 2010 ont été dérivées à partir de la méthode de triple collocation au chapitre 4. Cette étude a été menée à l'échelle globale puis continentale. Ces erreurs ont pu être comparées à d'autres paramètres physiques et algorithmiques : fraction vue par le radiomètre de forêt (FFO), de zone humide (FWL), d'eau pure (FWP), d'eau salée (FWS), de sol aride (FEB), de glace (FEI), de zone urbaine (FEU), probabilité d'interférence (RFI).

La méthode de classification CART (*Classification and Regression Tree*) a été utilisée pour définir des arbres de décision à l'échelle globale puis par continent. Cela permet de prédire une erreur moyenne en fonction d'un ensemble de valeurs de différents paramètres. Ainsi il est possible de mettre en évidence des ensembles de valeurs qui conduisent à de très fortes ou de très faibles erreurs SMOS.

L'arbre de décision à l'échelle globale est en figure 4.15, p. 80.

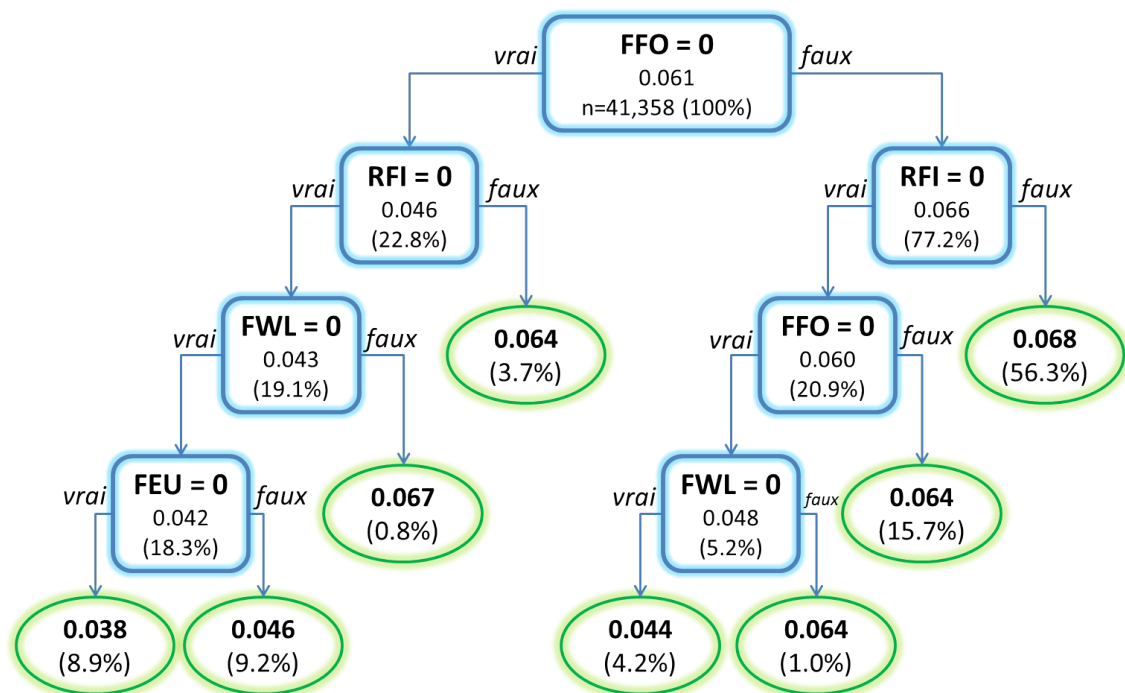


(a) Amérique du Nord

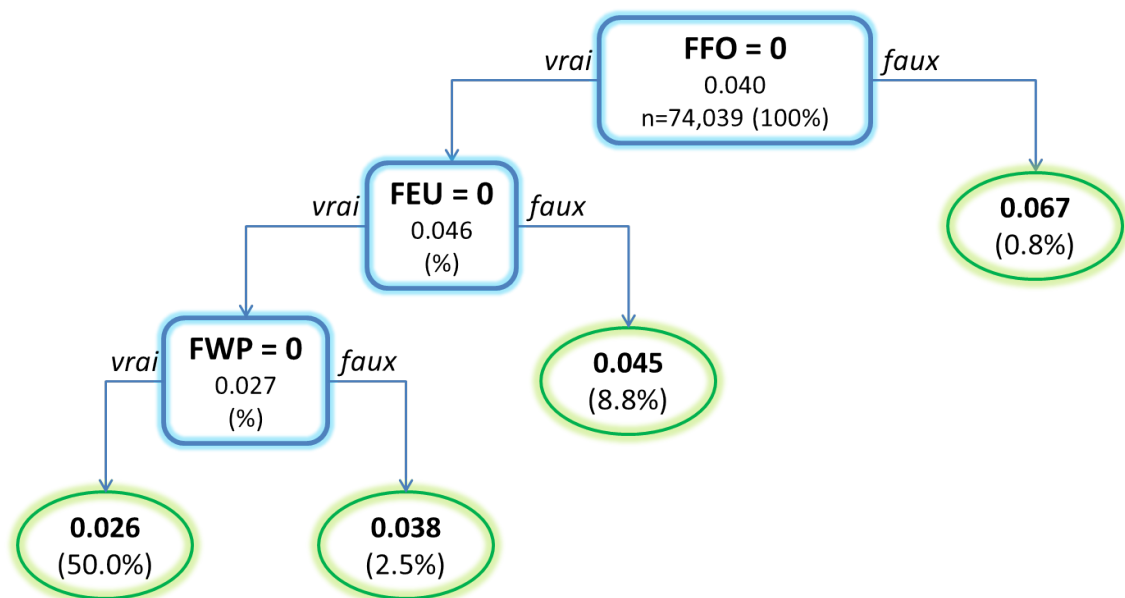


(b) Amérique du Sud

FIGURE B.1 – Arbre de régression de l'erreur SMOS à l'échelle continentale. Dans les rectangles sont inscrits les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre de points qui sont concernés. Si la condition est remplie, alors la branche *vraie* de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.

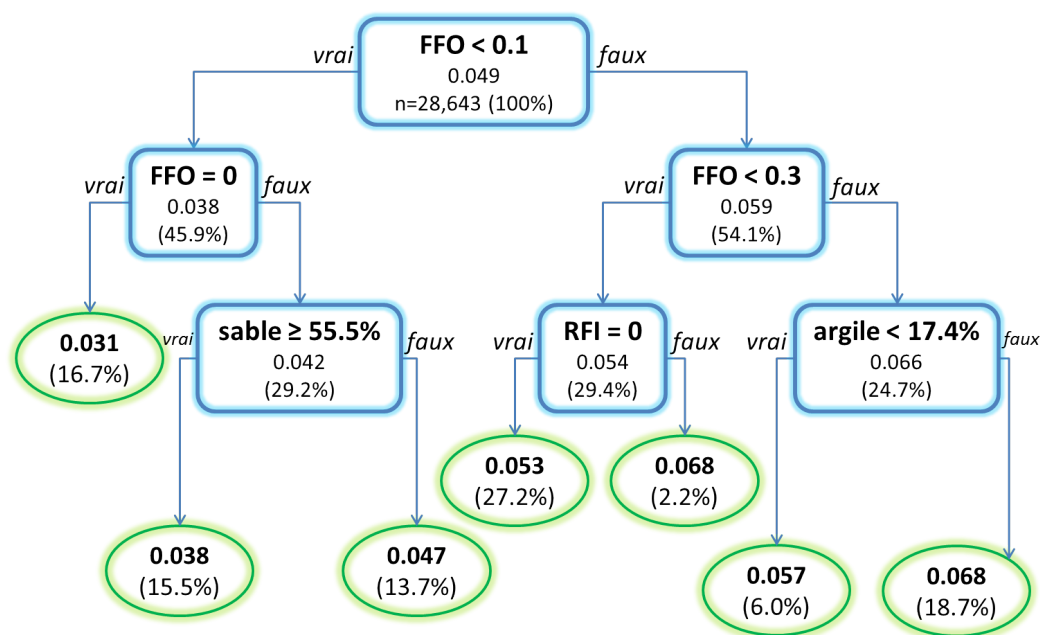


(c) Europe

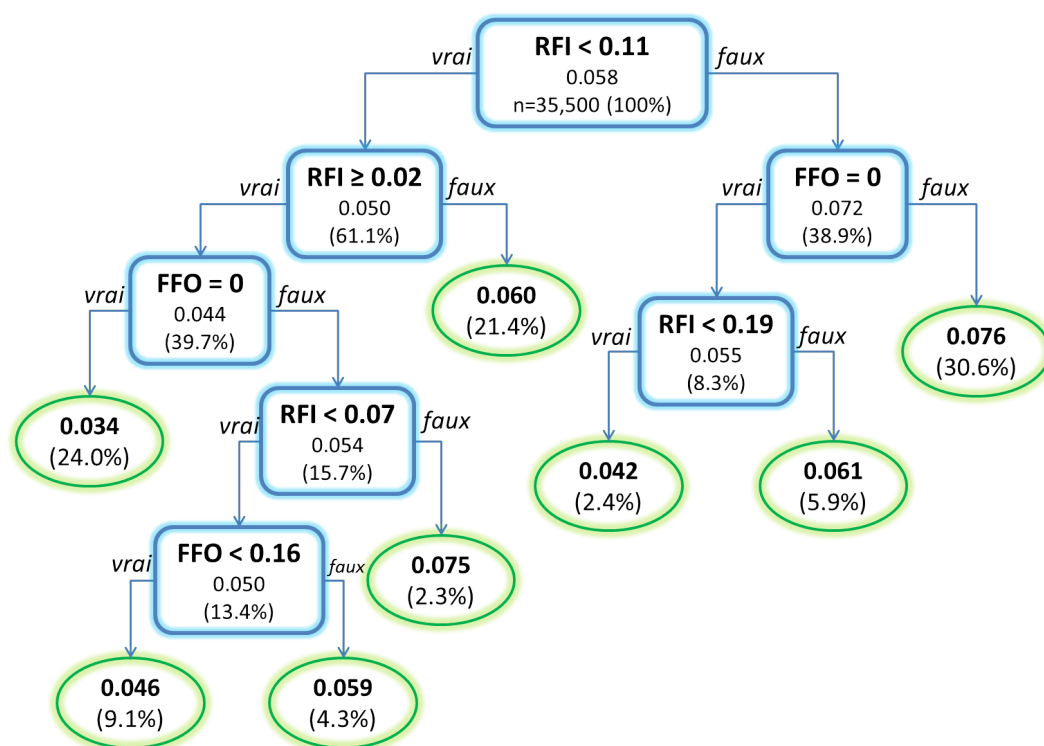


(d) Afrique du Nord

FIGURE B.1 – Arbre de régression de l'erreur SMOS à l'échelle continentale. Dans les rectangles sont inscrits les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre de points qui sont concernés. Si la condition est remplie, alors la branche *vraie* de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.

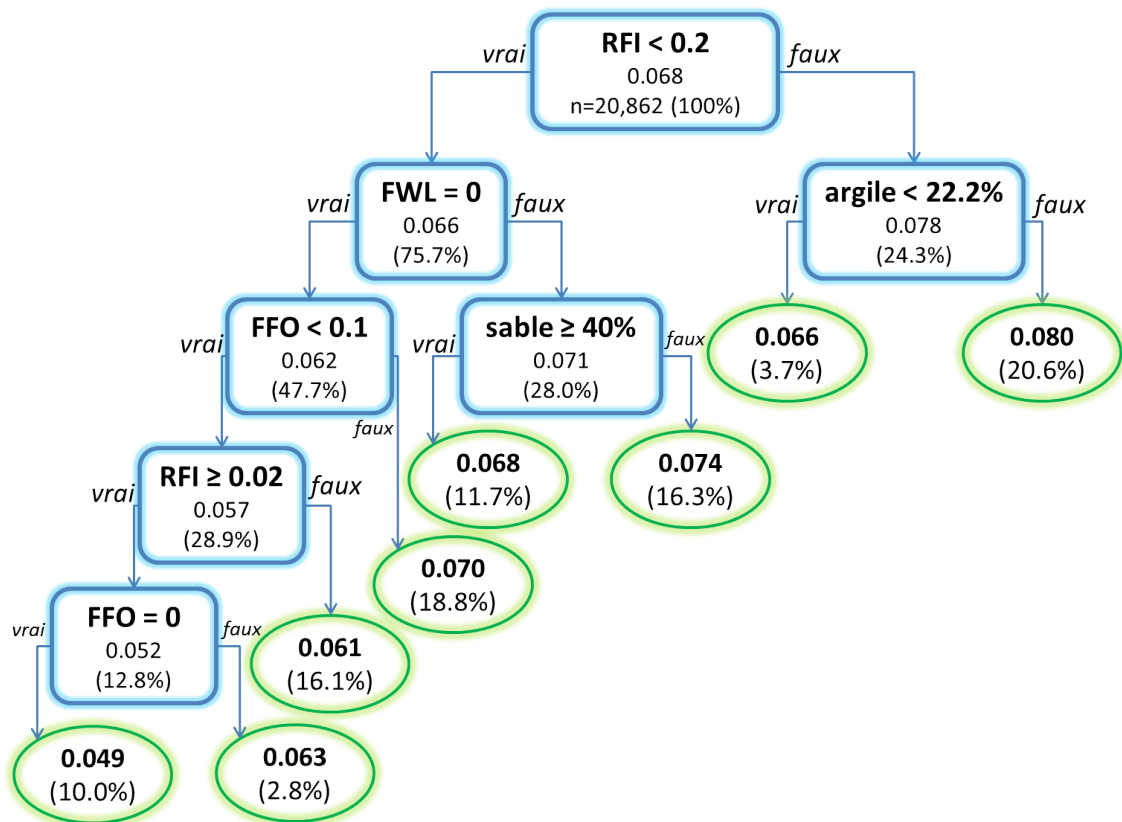


(e) Afrique du Sud

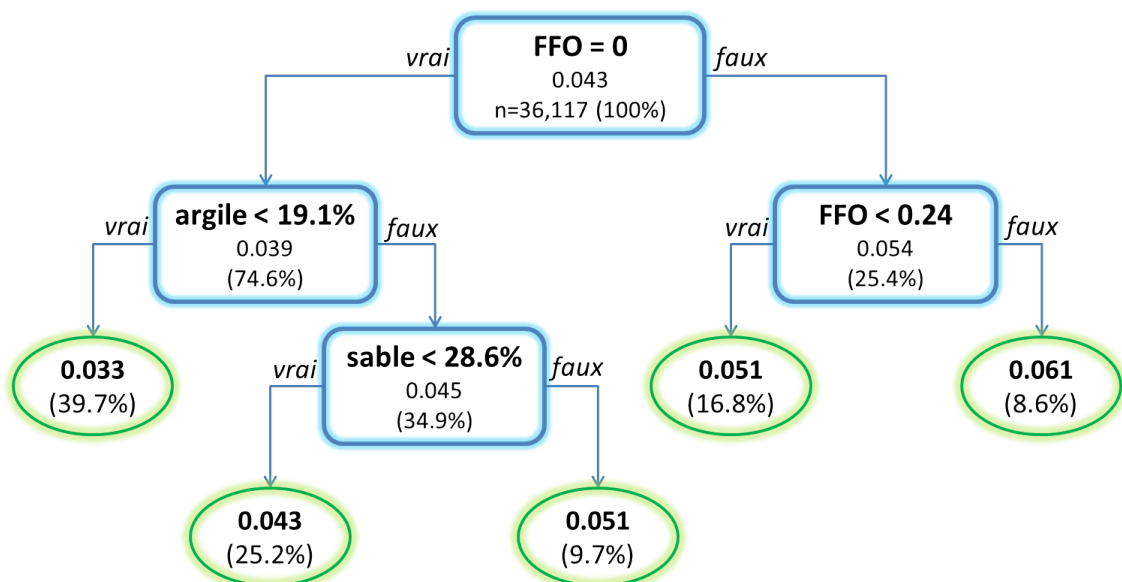


(f) Asie centrale

FIGURE B.1 – Arbre de régression de l'erreur SMOS à l'échelle continentale. Dans les rectangles sont inscrits les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre de points qui sont concernés. Si la condition est remplie, alors la branche *vraie* de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.



(g) Asie de l'Est



(h) Australie

FIGURE B.1 – Arbre de régression de l'erreur SMOS à l'échelle continentale. Dans les rectangles sont inscrits les conditions de découpage, ainsi que l'erreur moyenne de l'erreur avant la décision et le nombre de points qui sont concernés. Si la condition est remplie, alors la branche *vraie* de l'arbre est suivie. La fin de chaque branche se termine par une feuille représentée par un cercle avec la valeur moyenne de l'erreur SMOS (en m^3/m^3) et le nombre de points.

Modèles de constante diélectrique du sol

C.1 Modèle de Wang-Schmugge (Wang & Schmugge, 1980)

Pour la suite de cette section, voici la liste des paramètres utilisés :

k	: constante diélectrique du sol
SM	: humidité du sol [m^3/m^3]
P	: porosité du sol sec [-]
γ	: paramètre empirique [-]
SM_t	: humidité de transition [m^3/m^3]
k_a, k_w, k_r, k_i	: constantes diélectriques de l'air, de l'eau, de la roche et de la glace [-]
k_x	: constante diélectrique de l'eau absorbée initialement

La constante diélectrique k d'un mélange sol-eau est décrit par :

$$SM \leq SM_t \Rightarrow k = SM \cdot k_x + (P - SM) \cdot k_a + (1 - P) \cdot k_r$$

$$k_x = k_i + (k_w - k_i) \cdot SM / SM_t \cdot \gamma \quad (\text{C.1})$$

$$SM > SM_t \Rightarrow k = SM_t \cdot k_x + (SM - SM_t) \cdot k_w + (P - SM) \cdot k_a + (1 - P) \cdot k_r$$

$$k_x = k_i + (k_w - k_i) \cdot \gamma \quad (\text{C.2})$$

L'humidité de transition SM_t est définie comme la quantité d'humidité à laquelle la phase d'eau libre commence à dominer le mélange sol-eau :

$$SM_t = 0.49 \cdot WP + 0.165 \quad (\text{C.3})$$

avec WP le point de flétrissement, *wilting point* en anglais, en m^3/m^3 . Lorsque cette valeur n'est pas connue, elle peut être déterminée par :

$$WP = 0.06774 - 0.00064 \cdot SABLE + 0.004778 \cdot ARGILE \quad (\text{C.4})$$

avec $SABLE$ et $ARGILE$ les pourcentages de sable et d'argile contenus dans un sol sec.

Le paramètre empirique γ est défini par :

$$\gamma = -0.57 \cdot WP + 0.481 \quad (\text{C.5})$$

Les constantes diélectriques de l'air, de la roche et de la glace sont connues :

$$k_a = 1 \quad (\text{C.6})$$

$$k_r = 5.5 - 0.2i \quad (\text{C.7})$$

$$k_i = 3.2 - 0.1i \quad (\text{C.8})$$

La constante diélectrique de l'eau est donnée par l'équation de Debye :

$$k_w = k_{w\infty} + \frac{k_{w0} - k_{w\infty}}{1 + (2\pi t_w f)i} \quad (\text{C.9})$$

avec $k_{w\infty}$ la limite en fréquence haute de la constante diélectrique de l'eau pure (~ 4.9), k_{w0} la constante diélectrique statique de l'eau pure, t_w le temps de relaxation de l'eau pure en s et f la fréquence d'observation en Hz.

La constante diélectrique statique et le temps de relaxation de l'eau pure sont définis par :

$$k_{w0} = 88.045 - 0.4147 \cdot T + 6.295 \cdot 10^{-4} \cdot T^2 + 1.075 \cdot 10^{-5} \cdot T^3 \quad (\text{C.10})$$

$$2\pi t_w = 1.1109 \cdot 10^{-10} - 3.824 \cdot 10^{-12} \cdot T + 6.938 \cdot 10^{-14} \cdot T^2 - 5.096 \cdot 10^{-16} \cdot T^3 \quad (\text{C.11})$$

avec T la température effective de la couche du sol émettrice en $^{\circ}\text{C}$.

C.2 Modèle empirique de Dobson (Dobson et al., 1985) avec modification de (Peplinski et al., 1995) pour la bande L

Pour la suite de cette section, voici la liste des paramètres utilisés :

k	: constante diélectrique du sol
ρ	: densité du sol qui est fonction de sa texture $[\text{g}/\text{cm}^3]$
ρ_S	: densité des particules du sol, $\rho_S = 2.664 \text{ g}/\text{cm}^3$
k_{pa}	: constante diélectrique des particules du sol, $k_{pa} \approx 4.7$
α	: paramètre fixé à 0.65
β	: $\beta' - i\beta''$ est une fonction complexe dérivée empiriquement de la texture du sol, Peplinski et al. (1995)
SM	: humidité du sol $[\text{m}^3/\text{m}^3]$
k_w	: $k'_w - ik''_w$ est la constante diélectrique de l'eau libre dans le sol

(Dobson et al., 1985) proposent une expression pour la constante diélectrique d'un mélange sol-eau :

$$k = \left(1 + \frac{\rho}{\rho_S} (k_{pa}^{\alpha} - 1) + SM^{\beta'} \cdot k'_w{}^{\alpha} - SM \right)^{1/\alpha} - i \left(SM^{\beta''} \cdot k''_w{}^{\alpha} \right)^{1/\alpha} \quad (\text{C.12})$$

(Peplinski et al., 1995) ont exprimé les coefficients β' et β'' en fonction des fractions de sable (S) et d'argile (A) que contiennent le sol :

$$\beta' = 1.2748 - 0.519 \cdot S - 0.152 \cdot A \quad (\text{C.13})$$

$$\beta'' = 1.33797 - 0.603 \cdot S - 0.166 \cdot A \quad (\text{C.14})$$

La constante diélectrique de l'eau a été présentée dans la section précédente.

Le modèle de Dobson est un modèle de constant diélectrique très utilisé mais présente des faiblesses bien identifiées sur les sols secs et plus particulièrement sur les sols sableux. Les sols sableux ont tendance à être vus plus froids ce qui abaisse l'humidité du sol.

C.3 Modèle semi-empirique de Mironov (*Mironov et al., 2004*)

Le modèle de constante diélectrique de Mironov est basé sur l'indice de réfraction du sol. Il a besoin de moins d'entrées que le modèle de Dobson. Voici la liste des paramètres utilisés :

- k : constante diélectrique du sol
- η : indice de réfraction
- κ : coefficient d'atténuation normalisé

La constante diélectrique k est décomposée en partie réelle k' et partie imaginaire k'' :

$$k' = \eta^2 - \kappa^2 \tag{C.15}$$

$$k'' = 2\eta \cdot \kappa \tag{C.16}$$

L'indice de réfraction η et le coefficient de réfraction κ sont calculés à partir de nombreux paramètres dépendant de la fréquence, du contenu en argile, de l'humidité, etc. Les formules sont nombreuses et sont détaillées dans (*Mironov et al., 2004*).

Modélisations de la température efficace

D.1 Théorie

La température efficace est théoriquement une somme ou une intégration de la température du sol sur sa profondeur :

$$T_{eff} = \int_0^{\infty} T_s(z) \cdot W(z) \cdot dz \quad (D.1)$$

avec $T_s(z)$ la température thermodynamique du sol à la profondeur z et $W(z)$ la fonction de pondération qui est déterminée par :

$$W(z) = \alpha(z) \cdot \exp \left(- \int_0^z \alpha(z') \cdot dz' \right) \quad (D.2)$$

avec $\alpha(z)$ le coefficient d'atténuation défini par :

$$\alpha(z) = \frac{4\pi}{\lambda} \cdot \frac{k''(z)}{2\sqrt{k'(z)}} \quad (D.3)$$

avec λ la longueur d'onde d'observation, $k'(z)$ et $k''(z)$ les parties réelles et imaginaires de la constante diélectrique du sol à la profondeur z .

D.2 Modèle de Choudhury

Le modèle de Choudhury (Choudhury et al., 1982) ont proposé une paramétrisation de T_{eff} comme une fonction linéaire de la température du sol à deux profondeurs :

$$T_{eff} = T_{deep} - C \cdot (T_{deep} - T_{surf}) \quad (D.4)$$

avec T_{deep} la température du sol à une profondeur entre 50 et 100 cm, T_{surf} la température du sol à sa surface et C une constante dépendant de la longueur d'onde d'observation : $C = 0.802$ pour $\lambda = 2.8$ cm et $C = 0.084$ pour $\lambda = 49$ cm.

Choudhury et al. (1982) ont montré que la relation $T_{eff} = T_{surf}$ ($C=1$) n'est pas valide pour des fréquences inférieures à 10 GHz.

D.3 Modèle de Wigneron

Wigneron et al. (2001) proposent une amélioration en prenant en compte l'influence de l'humidité sur l'épaisseur de la couche de sol contribuant au signal : plus le sol est sec, plus profonde est la couche qui contribue.

$$T_{eff} = T_{deep} - C(SM_{surf}) \cdot (T_{deep} - T_{surf}) \quad (D.5)$$

avec T_{deep} la température du sol à une profondeur entre 50 et 100 cm, T_{surf} la température du sol à sa surface et C une constante dépendant de l'humidité de surface SM_{surf} :

$$C(SM_{surf}) = \left(\frac{SM_{surf}}{SM_0} \right)^b \quad (D.6)$$

avec SM_0 et b des paramètres ajustés pour chaque type de végétation.

D.4 Modèle de Holmes

Holmes et al. (2006) ont proposé une nouvelle paramétrisation de T_{eff} basée sur la relation entre l'atténuation de l'énergie émise et la constante diélectrique de la surface du sol :

$$T_{eff} = T_{deep} - C(k) \cdot (T_{deep} - T_{surf}) \quad (D.7)$$

avec T_{deep} la température du sol à une profondeur entre 50 et 100 cm, T_{surf} la température du sol à sa surface et C une constante dépendant de la constante diélectrique k du sol :

$$C(k) = \left(\frac{k''/k'}{k_0} \right)^b \quad (D.8)$$

avec k' et k'' les parties réelles et imaginaires de la constante diélectrique du sol, k_0 et b des paramètres ajustés pour chaque type de végétation.

Annexe E

Retrouver l'épaisseur optique de la végétation à partir des températures de brillance

Pour la suite de cette section, voici la liste des paramètres utilisés :

TB_V, TB_H	: températures de brillance en polarisation verticale ou horizontale [K]
T	: température du sol et de la végétation ($T_{eff} \approx T_{sol} \approx T_{vegetation}$) [K]
e_V, e_H	: émissivité du sol en polarisation verticale ou horizontale [-]
Γ	: transmissivité de la végétation [-]
ω	: albédo [-]
RP	: rapport de polarisation des températures de brillance [-]
τ	: épaisseur optique de la végétation [-]
θ	: angle d'incidence [rad]
k	: constante diélectrique du sol avec k' et k'' ses parties réelles et imaginaires.

(Meesters et al., 2005) ont proposé une formule analytique de l'épaisseur optique de la végétation à partir du rapport de polarisation des températures de brillance. Les équations du transfert radiatif qui sont la base de l'analyse ne tiennent pas compte des effets de l'atmosphère :

$$TB_V = T \cdot e_V \cdot \Gamma + (1 - \Gamma) \cdot T \cdot (1 - \omega) + (1 - e_V) \cdot (1 - \omega) \cdot T \cdot (1 - \Gamma) \cdot \Gamma \quad (E.1)$$

$$TB_H = T \cdot e_H \cdot \Gamma + (1 - \Gamma) \cdot T \cdot (1 - \omega) + (1 - e_H) \cdot (1 - \omega) \cdot T \cdot (1 - \Gamma) \cdot \Gamma \quad (E.2)$$

Le rapport de polarisation est défini comme suit :

$$RP = \frac{TB_V - TB_H}{TB_V + TB_H} \quad (E.3)$$

Si on injecte le modèle de transfert radiatif dans l'expression du rapport de polarisation, on obtient :

$$\frac{1}{RP} = \frac{e_V + e_H}{e_V - e_H} + \frac{2(1 - \omega)(1 - \Gamma^2)}{(\omega\Gamma + (1 - \omega)\Gamma^2)(e_V - e_H)} \quad (E.4)$$

Cette dernière équation revient en réalité à résoudre une équation du second degré :

$$\left[\frac{1}{2} \left(\frac{e_V - e_H}{PR} - (e_V + e_H) \right) + 1 \right] \Gamma^2 + \left[\frac{1}{2} \frac{\omega}{1 - \omega} \left(\frac{e_V - e_H}{PR} - (e_V + e_H) \right) \right] \Gamma - 1 = 0 \quad (E.5)$$

On pose :

$$a = \frac{1}{2} \left(\frac{e_V - e_H}{PR} - (e_V + e_H) \right) \quad (\text{E.6})$$

$$d = \frac{1}{2} \frac{\omega}{1 - \omega} \quad (\text{E.7})$$

En résolvant cette équation du second degré, une expression pour la transmissivité Γ de la végétation est obtenue :

$$\frac{1}{\Gamma} = a \cdot d + \sqrt{(a \cdot d)^2 + a + 1} \quad (\text{E.8})$$

L'épaisseur optique et la transmissivité de la végétation sont reliées par une expression exponentielle : $\Gamma = \exp(-\tau / \cos \theta)$ avec θ l'angle d'incidence. De ce fait on obtient l'expression finale de l'épaisseur optique :

$$\tau = \cos \theta \cdot \ln \left(a \cdot d + \sqrt{(a \cdot d)^2 + a + 1} \right) \quad (\text{E.9})$$

Annexe **F**

Étude du paramètre d'albédo optimisé avec
une rugosité dépendante ou non de
l'humidité du sol

F.1 Albédo optimisé

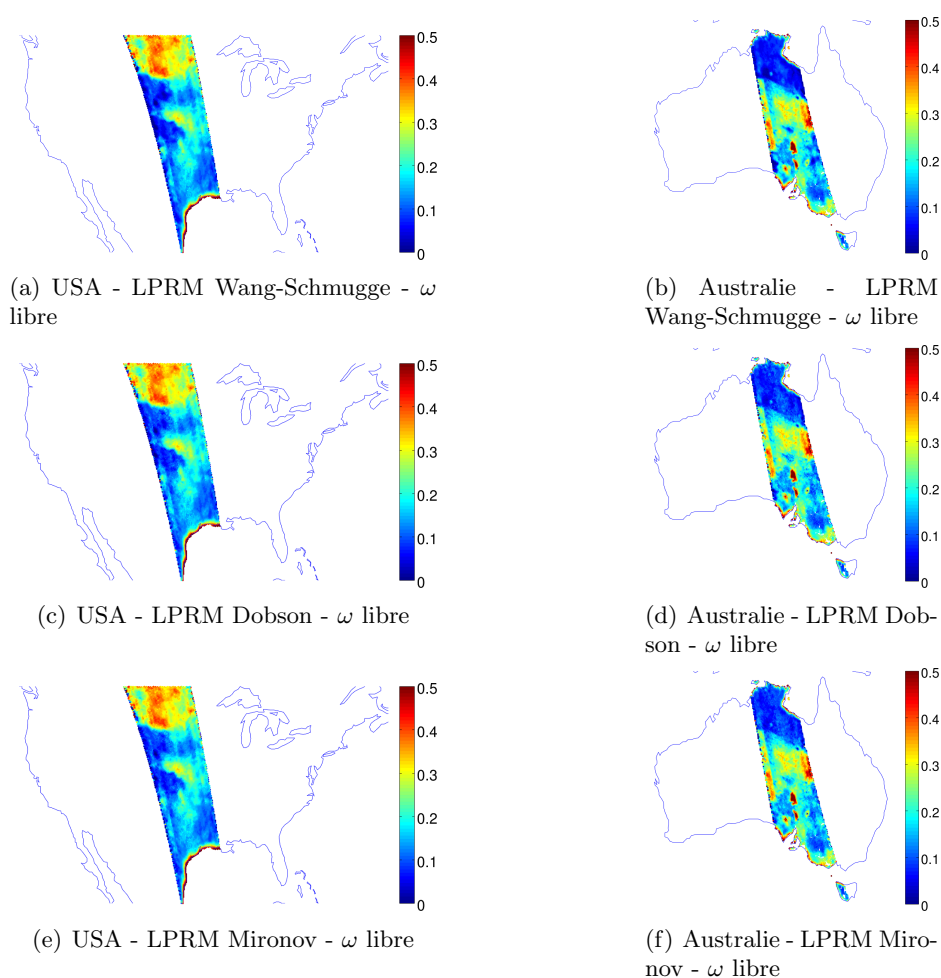


FIGURE F.1 – Représentations spatiales de l'albédo ω optimisée avec les trois modèles de constante diélectrique en libérant le paramètre d'albédo ω sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,b) LPRM avec le modèle de Wang-Schmugge, (c,d) LPRM avec le modèle de Dobson, (e,f) LPRM avec le modèle de Mironov.

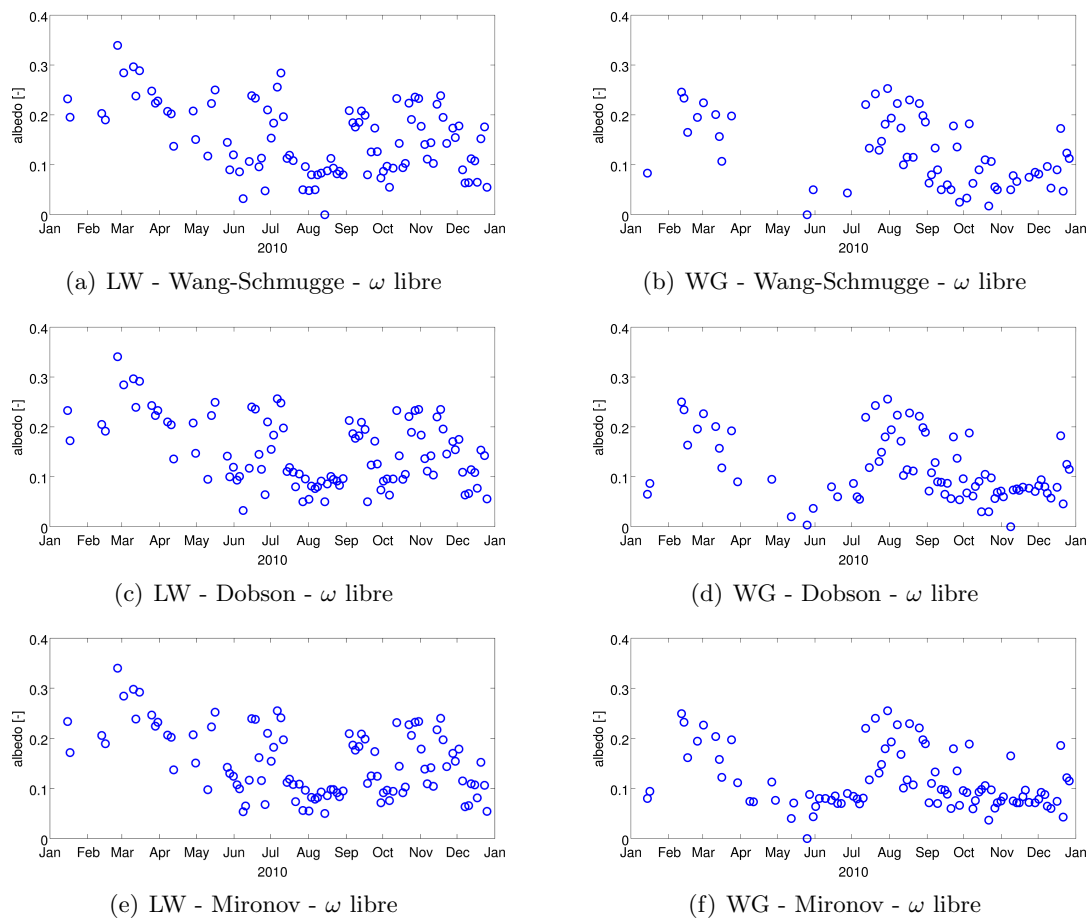


FIGURE F.2 – Représentations temporelles des albédos retrouvés avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.

F.2 Albédo optimisé avec la rugosité dépendante de l'humidité du sol

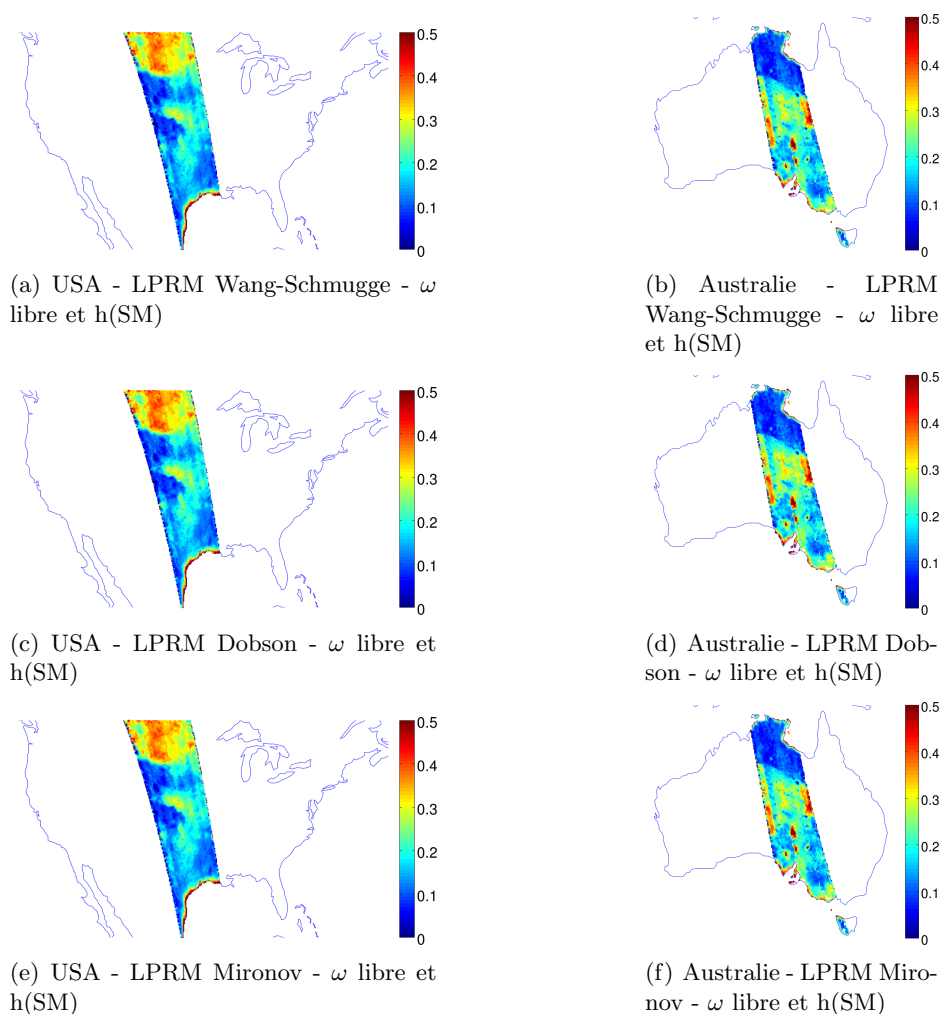


FIGURE F.3 – Représentations spatiales de l'albédo ω retrouvé à partir des trois modèles de constante diélectrique en libérant ω avec une dépendance de la rugosité à l'humidité du sol avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,b) LPRM avec le modèle de Wang-Schmugge, (c,d) LPRM avec le modèle de Dobson, (e,f) LPRM avec le modèle de Mironov.

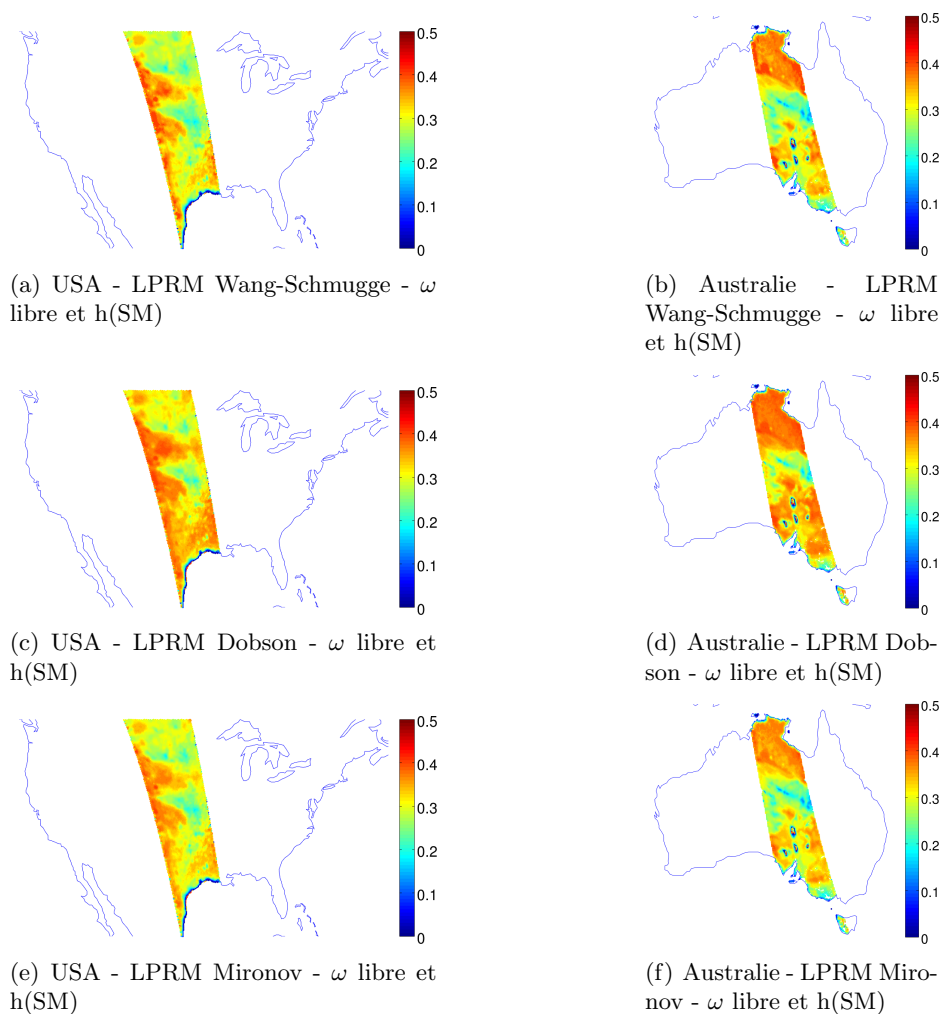
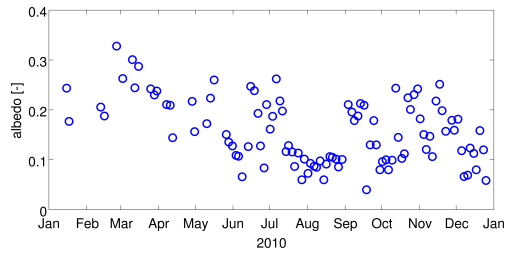
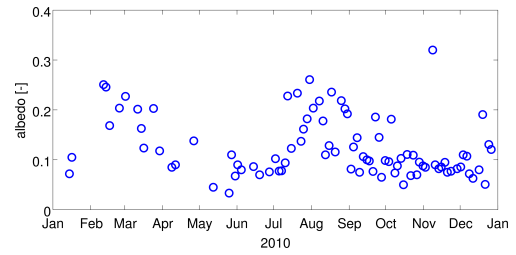


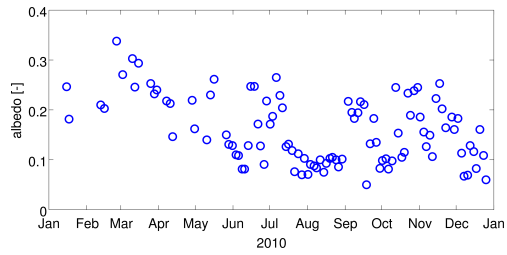
FIGURE F.4 – Représentations spatiales de la rugosité à 42.5° retrouvée à partir des trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol avec le produit d'humidité SMOS sur deux parties d'orbites aux États-Unis (16/09/2010) et en Australie (16/09/2010) : (a,b) LPRM avec le modèle de Wang-Schmugge, (c,d) LPRM avec le modèle de Dobson, (e,f) LPRM avec le modèle de Mironov.



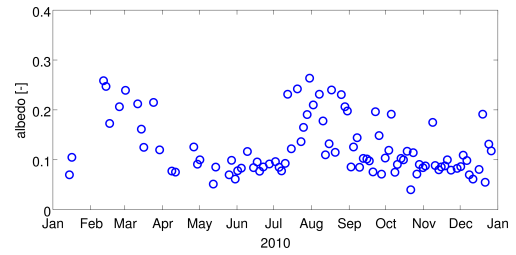
(a) LW - Wang-Schmugge - ω libre et h(SM)



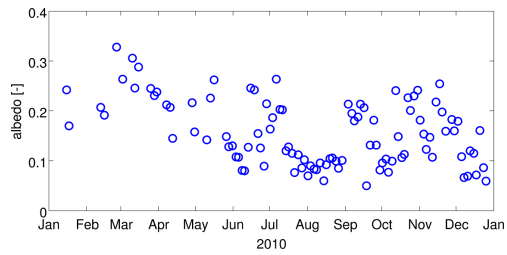
(b) WG - Wang-Schmugge - ω libre et h(SM)



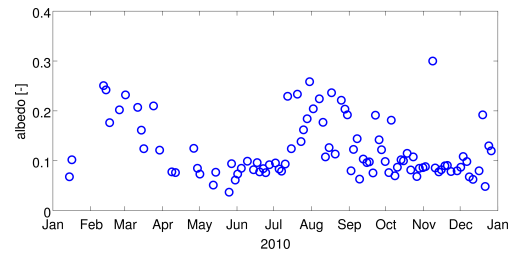
(c) LW - Dobson - ω libre et h(SM)



(d) WG - Dobson - ω libre et h(SM)



(e) LW - Mironov - ω libre et h(SM)



(f) WG - Mironov - ω libre et h(SM)

FIGURE F.5 – Représentations temporelles des albédos retrouvés avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.

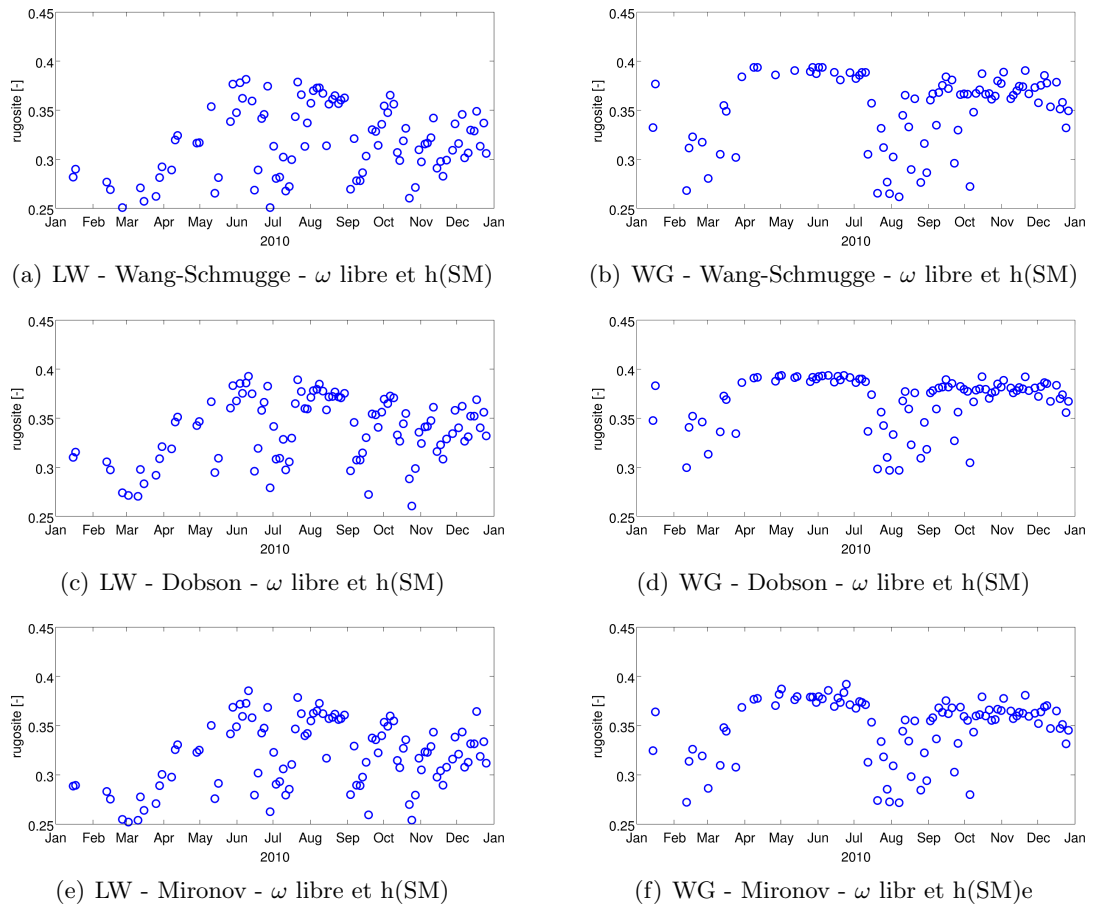


FIGURE F.6 – Représentations temporelles des rugosités à 42.5° retrouvées avec le LPRM et trois modèles de constante diélectrique en libérant le paramètre d'albédo ω avec une dépendance de la rugosité à l'humidité du sol sur Little Washita (LW) et Walnut Gulch (WG) en 2010 : (a,b) Wang-Schmugge, (c,d) Dobson et (e,f) Mironov.

Historique des versions du produit d'humidité SMOS de niveau 2

Les informations concernant les changements de version proviennent du site internet de la compagnie Array qui développe les processeurs de l'algorithme de niveau 2 SMOS de façon opérationnelle. Ces informations peuvent être trouvées à l'adresse suivante :

[http://smos.array.ca/c/document_library/get_file?uuid=5ab03a16-c226-49e2-9d2d-](http://smos.array.ca/c/document_library/get_file?uuid=5ab03a16-c226-49e2-9d2d-80fef6a9614e&groupId=11352)

[80fef6a9614e&groupId=11352](http://smos.array.ca/c/document_library/get_file?uuid=5ab03a16-c226-49e2-9d2d-80fef6a9614e&groupId=11352) .

L2SM processor version	Updates	Deployment at DPGS
v3.05	- Commissioning version	Jan 28, 2010
v3.06	- Commissioning version	20100214T201707
v3.07	- Commissioning version	20100303T082453
v3.08	- New convention in L1c for Faraday rotation, from surface to antenna - Resetting of retrieved geophysical parameters and associated DQX to -999 when retrieval fails - Correction for the computation of forest Tau under snow condition	20100707T045941
v3.09	- Updated code to deal with a bug related to DQX=0 in AUX_DGGTFO, AUX_DGGTLV, AUX_DGGROU - C_RFI adjusted so its value is always ≤ 1 - Corrected reporting of the causes of the retrieval failures under certain conditions - Fixed minor inconsistencies between SMUDP and SMDAP - Updated code to improve the reporting and initialization of several values in the SMUDP and SMDAP	20100809T060522

L2SM processor version	Updates	Deployment at DPGS
v4.00	<ul style="list-style-type: none"> - Updated ADFs (AUX_DFFLMX, AUX_CNFSMx, AUX_LANDCL) - Added pseudo dual retrieval capability, or the so called "dual-in-full" allowing the processor to filter out cross polarization data from the retrieval process due to the present errors - Implemented a new formulation for forest retrievals - Updated formulation of the quality index GQX to use Chi_2 instead of Chi_2_P - Made several enhancements to SMDAP to aid analysis - Suppressing the so called "rich retrievals" - Enabling retrievals of Tau over areas with very low vegetation - Updated the processor due to changes in AUX_ECMWF to allow scaling of SM parameter through CDF matching coefficients - Added EE CFI 3.7.3 and support for IERS Bulletin B - Included L1 RFI flag summaries in SMUDP - Fixed a bug related to Cardiod retrieval over water - Made improvements in the source and binary distribution of the code 	Never deployed, superseded by v4.01
v4.01	<ul style="list-style-type: none"> - Fixed a bug due to "poorly conditioned matrix" leading to segmentation fault - DSR_Size in SMUDP header corrected from 200 to 220 	20110517T211748
v5.00	<ul style="list-style-type: none"> - Aligned with the Post Processor (data-driven vs time driven) - Dynamic RFI detection based on target surface temperature/emissivity - RFI probability per DGG node reported in the SMUDP - Retrieves SM over moderate topography - Increased acceptance range of SM to allow measurement of extreme cases 	20111019T140444

L2SM processor version	Updates	Deployment at DPGS
v5.01	<ul style="list-style-type: none"> - Functionally equivalent to v5.00 but adds multi-thread capability with number of threads being a configurable parameter - Processing time significantly reduced even when a single thread is used - Fixed a bug where residuals in SMDAP did not match SMUDP values in case of retrieval failures 	Never deployed but used for 2011 reprocessing campaign
v5.50	<ul style="list-style-type: none"> - Added the capability to use Mironov dielectric model in addition to Dobson and allows selection of dielectric model using two switches contained in AUX_CNFSMx <ul style="list-style-type: none"> – Dielectric_Model_Type = 0/1 (Dobson/Mironov) – Dielectric_Model_Sub_Type = 0/1 (Standard/Symmetric) <p>where (Dobson/Standard) is equivalent to the operation of L2SM v5.00</p>	Never deployed, superseded by v5.51
v5.51	Fixed a bug causing failure at latitudes outside $[-89^{\circ}, +89^{\circ}]$	20120423T083517

Afin de distribuer des produits d'humidité cohérents, plusieurs campagnes de retraitement des données ont été effectuées. La dernière disponible permet d'avoir des humidités en v5.51 depuis le 10 Janvier 2010 jusqu'à nos jours.

Bibliographie

- Ahmed, N. (1995). Estimating soil moisture from 6.6 GHz dual polarization and/or satellite derived vegetation index. *International Journal of Remote Sensing*, 16, 687–708.
- Al Bitar, A., Leroux, D., Kerr, Y., Merlin, O., Richaume, P., Sahoo, A., & Wood, E. (2012). Evaluation of SMOS soil moisture products over continental US using SCAN/SNOTEL network. *IEEE Transaction on Geoscience and Remote Sensing*, 50, 1572–1586.
- Albergel, C., Rüdiger, C., Carrer, D., Calvet, J.-C., Fritz, N., Naeimi, V., Bartalis, Z., & Hasenauer, S. (). An evaluation of ascats surface soil moisture products with in-situ observations in southwestern france. *Hydrology and Earth System Sciences*, volume =, .
- Anterrieu, E., & Khazaal, A. (2008). Brightness temperature map reconstruction from dual-polarimetric visibilities in synthetic aperture imaging radiometry. *IEEE Transaction on Geoscience and Remote Sensing*, 46, 606–612.
- Armstrong, M. (). *Copula catalogue, part 1 : bivariate Archimedean copulas*, institution =. Technical Report.
- Arnold, H. (2006). *Dependence modelling via the copula method*. Technical Report CSIRO Mathematical and Information Sciences - Quantitative Risk Management Group.
- ATBD (2010). *Algorithm Theoretical Basis Document*. Technical Report 3.d.
- Australian Government (2006). <http://www.cazr.csiro.au/connect/resources.htm>, .
- Barrett, E., Kidd, C., & Bailey, J. (1988). The Special Sensor Microwave Imager - a new instrument with rainfall monitoring potential. *International Journal of Remote Sensing*, 9, 1943–1950.
- Basist, A., Grody, N., Peterson, T., & Williams, C. (1998). Using the special sensor microwave imager to monitor land surface temperature, wetness, and snow cover. *Journal of Applied Meteorology*, 37, 888–911.
- Becker, F., & Choudhury, B. (1988). Relative sensitivity of normalized difference vegetation index (NDVI) and microwave polarization difference index (MPDI) for vegetation and desertification monitoring. *Remote Sensing of Environment*, 24, 297–311.
- Begert, M., Schlegel, T., & Kirchhofer, W. (2005). Homogeneous temperature and precipitation series of switzerland from 1864 to 2000. *International Journal of Climatology*, 25, 65–80.

- Berg, D. (2009). Copula goodness-of-fit testing : an overview and power comparison. *European Journal of Finance*, 15, 675–701.
- Bioinformatics (2002). <http://www.bioinformatics.org/oeil-couleur/dossier/lumiere.html>, .
- Bircher, S., Balling, J., Skou, N., & Kerr, Y. (2012). Validation of SMOS brightness temperatures during the HOBE airborne campaign, Western Denmark. *IEEE Transaction on Geoscience and Remote Sensing*, 50, 1468–1482.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth Adv. Book Prog.
- Brock, F., Crawford, K., Elliott, R., Cuperus, R., Stadler, S., Johnson, H., & Eilts, M. (1995). The oklahoma mesonet : a technical overview. *Journal of Atmospheric and Oceanic Technology*, 12, 5–19.
- Brunfeldt, D., & Ulaby, F. (1984). Measured microwave emission and scattering in vegetation canopies. *IEEE Transaction on Geoscience and Remote Sensing*, GE-22, 520–524.
- Caires, S., & Sterl, A. (2003). Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical research*, 108, 43(1)–43(15).
- Chaurasia, S., Tung, D., Thapliyal, P., & Joshi, P. (2011). Assessment of AMSR-E soil moisture product over India. *International Journal of Remote Sensing*, 32, 7955–7970.
- Chen, J., & Black, T. A. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell and Environment*, 15, 421–429.
- Choi, M., & Jacobs, J. (2008). Temporal variability corrections for Advanced Microwave Scanning Radiometer E (AMSR-E) surface soil moisture : case study in Little River region, Georgia, U.S. *Sensors*, 8, 2617–2627.
- Choudhury, B., Owe, M., Goward, S., Golus, R., Ormsby, J., Chang, A., & Wang, J. (1987). Quantifying spatial and temporal variabilities of microwave brightness temperature over the U.S. southern great plains. *International Journal of Remote Sensing*, 8, 177–191.
- Choudhury, B., Schmugge, T., Chang, A., & Newton, R. (1979). Effect of surface roughness on the microwave emission from soils. *Journal of Geophysical Research*, 84, 5699–5705.
- Choudhury, B., Schmugge, T., & Mo, T. (1982). A parameterization of effective soil temperature for microwave emission. *Journal of Geophysical Research*, 87, 1301–1304.
- Cosh, M., Jackson, T., Bindlish, R., Famiglietti, J., & Ryu, D. (2005). Calibration of an impedance probe for estimation of surface soil water content over large regions. *Journal of Hydrology*, 311, 49–58.
- Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., & Elrayes, M. A. (1985). Microwave dielectric behavior of wet soil .2. dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing*, 23, 35–46.
- Dorigo, W., Scipal, K., Parinussa, R., Liu, Y., Wagner, W., de Jeu, R., & Naeimi, V. (2010). Error characterisation of global active and passive microwave soil moisture data sets. *Hydrology and Earth System Sciences*, 7, 5621–5645.
- Douville, H., & Chauvin, F. (2000). Relevance of soil moisture for seasonal climate predictions : a preliminary study. *Climate Dynamics*, 16, 719–736.

- Draper, C., Walker, J., Steinle, P., de Jeu, R., & Holmes, T. (2009). An evaluation of AMSR-E derived soil moisture over Australia. *Remote Sensing of Environment*, 113, 703–710.
- Drusch, M. (2007). Initializing numerical weather predictions models with satellite derived surface soil moisture : data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set. *Journal of Geophysical Research*, 113.
- Drusch, M., Wood, E., & Gao, H. (2005). Observation operators for the direct assimilation of TRMM Microwave Imager retrieved soil moisture. *Geophysical Research Letters*, 32.
- Dupuis, D. (2007). Using copulas in hydrology : benefits, cautions and issues. *Journal of Hydrologic Engineering*, 12, 381–393.
- Escorihuela, M., Kerr, Y., de Rosnay, P., Wigneron, J., Calvet, J., & Lemaitre, F. (2007). Simple model of the bare soil microwave emission at l-band. *IEEE Transactions on Geoscience and Remote Sensing*, 45, 1978–1987.
- Evelt, S., Tolk, J., & Howell, T. (2006). Soil profile water content determination : sensor accuracy, axial response, calibration, temperature dependence and precision. *Soil Science Society of America*, 5, 894–907.
- Favre, A., Adlouni, S. E., Perreault, L., Thiémonge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40.
- Fermanian, J. (2005). Goodness of fit tests for copulas. *Journal of Multivariate Analysis*, 95, 119–152.
- Ferrazzoli, P., & Guerriero, L. (1996). Passive microwave remote sensing of forests : A model investigation. *IEEE Transactions on Geoscience and Remote Sensing*, 34, 433–443.
- Figa-Saldana, J., Wilson, J., Attema, E., Gelsthorpe, R., Drinkwater, M., & Stoffelen, A. (2002). The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform : A follow on for European wind scatterometers. *Canadian Journal of Remote Sensing*, 28, 404–412.
- Frees, E., & Valdez, E. (1998). Understanding relationships using copulas. *North America Actuarial Journal*, 2, 1–25.
- Gaiser, P., St-Germain, K., Twarog, E., Poe, G., Purdy, W., Richardson, D., Grossman, W., Jones, W., Spencer, D., Golba, G., Cleveland, J., Choy, L., Bevilacqua, R., & Chang, P. (2004). The WindSat spaceborne polarimetric microwave radiometer : Sensor description and early orbit performance. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 2347–2361.
- Gao, H., Wood, E., Drusch, M., & McCabe, M. (2007). Copula-derived observation operators for assimilating TMI and AMSR-E retrieved soil moisture into land surface models. *Journal of Hydrometeorology*, (pp. 413–429).
- Genest, C., & Favre, A. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12, 347–368.
- Genest, C., Quessy, J.-F., & Remillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transform. *Scandinavian Journal of Statistics*, 33, 337–366.

- Genest, C., & Rivest, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88, 1034–1043.
- Genest, C., & Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales Henri Poincaré*, 44.
- Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas : a review and a power study. *Insurance : Mathematics and Economics*, 44, 199–213.
- Gherboudj, I., Magagi, R., Goita, K., Berg, A., Toth, B., & Walker, A. (2012). Validation of SMOS data over agricultural and boreal forest areas in Canada. *IEEE Transaction on Geoscience and Remote Sensing*, 50, 1623–1635.
- Gini, C. (1912). *Italian : Variabilità e mutabilità (Variability and Mutability)*. C. Cuppini.
- Gruhier, C., de Rosnay, P., Hasenauer, S., Holmes, T., de Jeu, R., Kerr, Y., Mougin, E., Njoku, E., Timouk, F., Wagner, W., & Zribi, M. (2010). Soil moisture active and passive microwave products : intercomparison and evaluation over a sahelian site. *Hydrology and Earth System Sciences*, 14, 141–156.
- Gruhier, C., de Rosnay, P., Kerr, Y., Mougin, E., Ceschia, E., Calvet, J., & Richaume, P. (2008). Evaluation of AMSR-E soil moisture product based on ground measurements over temperate and semi-arid regions. *Geophysical Research Letters*, 35.
- Guilfen, Y., Chapron, B., & Vandemark, D. (2001). The ERS scatterometer wind measurement accuracy : evidence of seasonal and regional biases. *Journal of Atmospheric and Oceanic Technology*, 18, 1684–1697.
- Hafner, C., & Reznikova, O. (2010). Efficient estimation of a semiparametric dynamic copula model. *Computational Statistics and Data Analysis*, 54, 2609–2627.
- Holmes, T., de Rosnay, P., de Jeu, R., Wigneron, J., Kerr, Y., Calvet, J., Escorihuela, M., Saleh, K., & Lemaitre, F. (2006). A new parameterization of the effective temperature for l band radiometry. *Geophysical Research Letters*, 23, L07405.
- Hornbuckle, B. (2012). communication personnelle.
- Huard, D., Evin, G., & Favre, A. (2006). Bayesian copula selection. *Computational Statistics and Data Analysis*, 51, 809–822.
- IPCC - Intergovernmental Panel on Climate Change (2007). *Climate Change 2007 : Synthesis Report*. Technical Report Core Writing Team, Pachauri, R.K. and Reisinger, A. (Eds.).
- Jackson, T., Bindlish, R., Cosh, M., Zhao, T., Starks, P., Bosch, D., Seyfried, M., Moran, S., Kerr, Y., & Leroux, D. (2012). Validation of Soil Moisture Ocean Salinity (SMOS) soil moisture over watershed networks in the U.S. *IEEE Transactions on Geoscience and Remote Sensing*, 50, 1530–1543.
- Jackson, T., Cosh, M., Bindlish, R., Starks, P., Bosch, D., Seyfried, M., Goodrich, D., Moran, M., & Du, J. (2010). Validation of Advanced Microwave Scanning Radiometer soil moisture products. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 4256–4272.
- Jackson, T., & O'Neill, P. (1990). Attenuation of soil microwave emission by corn and soybeans at 1.4 and 5 GHz. *IEEE Transactions on Geoscience and Remote Sensing*, 28, 978–980.

- Jackson, T., Schmugge, T., & Wang, J. (1982). Passive microwave sensing of soil moisture under vegetation canopies. *Water Resources Research*, 18, 1137–1142.
- Janssen, P., Abdalla, S., Hersbach, H., & Bidlot, J.-R. (2007). Error estimation of buoy, satellite and model wave height data. *Journal of Atmospheric and oceanic technology*, 24, 1665–1677.
- de Jeu, R., Holmes, T., Panciera, R., & Walker, J. (2009). Parameterization of the Land Parameter Retrieval Model for L-band observations using the NAFE'05 data set. *IEEE Geoscience and Remote Sensing Letters*, 6, 630–634.
- Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H., & Spencer, R. (2003). The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 184–194.
- Kerr, Y. (1996). *Optimal choice for MIRAS frequencies scientific requirements*. Technical Report CESBIO.
- Kerr, Y., & Njoku, E. (1990). A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space. *IEEE Transactions on Geoscience and Remote Sensing*, 28, 384–393.
- Kerr, Y., Waldteufel, P., Richaume, P., Wigneron, J., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S., Leroux, D., Mialon, A., & Delwart, S. (2012). The SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 50, 1384–1403.
- Kerr, Y., Waldteufel, P., Wigneron, J., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M., Font, J., Reul, N., Gruhier, C., Juglea, S., Drinkwater, M., Hahne, A., Martin-Neira, M., & Mecklenburg, S. (2010). The SMOS Mission : New Tool for Monitoring Key Elements of the Global Water Cycle. *Proceedings of the IEEE*, 98, 666–687.
- Kerr, Y., Waldteufel, P., Wigneron, J., Martinuzzi, J., Font, J., & Berger, M. (2001). Soil moisture retrieval from space : the Soil Moisture and Ocean Salinity (SMOS) Mission. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1729–1735.
- Kirdiashev, K., Chukhantsev, A., & Shutko, A. (1979). Microwave radiation of the Earth's surface in the presence of vegetation cover. *Radio Engineering and Electronic Physics*, 24, 256–264.
- Laux, P., Vogl, S., Qiu, W., Knoche, H., & Kunstmann, H. (2011). Copula-based statistical refinement of precipitation in rcm simulations over complex terrain. *Hydrology and Earth System Sciences*, 15, 2401–2419.
- Levaray, G. (2004). <http://fertifondp.siaap.fr/index.php?p=26>, .
- Li, H., Sheffield, J., & Wood, E. (2010a). Bias correction of monthly precipitation and temperature fields from IPCC AR4 models using equidistant quantile matching. *Journal of Geophysical Research-Atmospheres*, 115.
- Li, L., Gaiser, P., Gao, B., Bevilacqua, R., Jackson, T., Njoku, E., Rüdiger, C., Calvet, J., & Bindlish, R. (2010b). WindSat global soil moisture retrieval and validation. *IEEE Transactions on Geoscience and Remote Sensing*, 48.

- Liu, Y., van Dijk, A., Jeu, R. D., & Holmes, T. (2009). An analysis of spatiotemporal variations of soil and vegetation moisture from a 29-year satellite-derived data set over mainland Australia. *Water Resources Research*, 45.
- Liu, Y., Parinussa, R., Dorigo, W., Jeu, R. D., Wagner, W., Dijk, A. V., McCabe, M., & Evans, J. (2011). Developping an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences*, 15, 425–436.
- Masson, V., Champeau, J.-L., Chauvin, F., Meriguet, C., & Lacaze, R. (2003). A global data base of land surface parameters at 1 km resolution in meteorological and climate models. *Journal of Climate*, 16, 1261–1282.
- Meesters, A., de Jeu, R., & Owe, M. (2005). Analytical derivation of the vegetation optical depth from the microwave polarization difference index. *IEEE Transactions on Geoscience and Remote Sensing*, 2, 121–123.
- Mialon, A., Coret, L., Kerr, Y. H., Secherre, F., & Wigneron, J. P. (2008). Flagging the topographic impact on the smos signal. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 689–694.
- de Michele, C., & Salvadori, G. (2003). A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research*, 108.
- Miralles, D., Crow, W., & Cosh, M. (2010). Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *Journal of Hydrometeorology*, 11, 1423–1429.
- Mironov, V., Dobson, M., Kaupp, V., Komarov, S., & Kleshchenko, V. (2004). Generalized refractive mixing dielectric model for moist soils. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 773–785.
- Mironov, V., & Fomin, S. (2009). Temperature and mineralogy dependable model for microwave dielectric spectra of moist soils. *PIERS Online*, 5, 411–415.
- Mo, T., Choudhury, B., Wang, J., & Jackson, T. (1982). A model for microwave emission from vegetation-covered fields. *Journal of Geophysical Research*, 87, 11229–11237.
- Naeimi, V., Scipal, K., Bartalis, Z., & Wagner, S. H. W. (2009). An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 1999–2013.
- NASA (2008). http://fr.wikipedia.org/wiki/Fichier:Atmospheric_electromagnetic_opacity.svg, .
- Nelsen, R. (1998). *An introduction to copulas*. Springer Series in Statistics.
- Njoku, E. (2004). updated daily. AMSR-E/Aqua Daily L3 Surface Soil Moisture, Interpretive Parameters, QC EASE-Grids. Digital Media.
- Njoku, E., Jackson, T., Lakshmi, V., Chan, T., & Nghiem, S. (2003). Soil moisture retrieval from ASMR-E. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 215–229.
- Njoku, E., & Li, L. (1999). Retrieval of land surface parameters using passive microwave measurements at 6-18 GHz. *IEEE Transactions on Geoscience and Remote Sensing*, 37, 79–93.

- Njoku, E., Stacey, J., & Barath, F. (1980). The Seasat Scanning Multichannel Microwave Radiometer (SMMR) : instrument description and performance. *IEEE Journal of Oceanic Engineering*, 5, 100–115.
- O'Neill, P., Entekhabi, D., Njoku, E., & Kellogg, K. (2010). The NASA soil moisture active passive (SMAP) mission : overview. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, .
- Outwater, A. (1997). *Water, a natural history*. Basic Books.
- Owe, M., de Jeu, R., & Walker, J. (2001). A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1643–1654.
- Owe, M., de Griend, A. V., & Chang, A. (1992). Surface moisture and satellite microwave observations in semiarid southern Africa. *Water Resources Research*, 28, 829–839.
- Owe, M., & Van de Griend, A. (2001). On the relationship between thermodynamic surface temperature and high frequency (37 GHz) vertical polarization brightness temperature under semi-arid conditions. *International Journal of Remote Sensing*, 22, 3521–3532.
- Pampaloni, P., & Paloscia, S. (1986). Microwave emission and plant water content : a comparison between field measurements and theory. *IEEE Transactions on Geoscience and Remote Sensing*, 24, 900–905.
- Panciera, R., Walker, J., Kalma, J., Kim, E., Hacker, J., Merlin, O., Berger, M., & Skou, N. (2008). The NAFE'05/CoSMOS data set : toward SMOS soil moisture retrieval, downscaling, and assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 736–745.
- Panciera, R., Walker, J., Kalma, J., Kim, E., Saleh, K., & Wigneron, J. (2009). Evaluation of the SMOS L-MEB passive microwave soil moisture retrieval algorithm. *Remote Sensing of Environment*, 113, 435–444.
- Parrens, M., Zakharova, E., Lafont, S., Calvet, J.-C., Kerr, Y., Wagner, W., & Wigneron, J.-P. (2011). Comparing soil moisture retrievals from SMOS and ASCAT over France. *Hydrology and Earth System Sciences Discussions*, 8, 8565–8607.
- Peischl, S., Walker, J., Allahmoradi, M., Barrett, D., Gurney, R., Kerr, Y., Kim, E., LeMarshall, J., Rudiger, C., Ryu, D., & Ye, N. (2009). Towards validation of SMOS using airborne and ground data over the Murrumbidgee Catchment. *18th world IMACS congress and MODISM09 international congress on modelling and simulation : interfacing modelling and simulation with mathematical and computational sciences*, (pp. 3733–3739).
- Peplinski, N., Ulaby, F., & Dobson, M. (1995). Dielectric properties of soils in the 0.3-1.3 GHz range. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 803–807.
- Picard, G., & Fily, M. (2006). Surface melting observations in antarctica by microwave radiometers : correcting 26-year time series from changes in acquisition hours. *Remote Sensing of Environment*, 104, 325–336.
- Prigent, C., Papa, F., Aires, F., Jimenez, C., Rossow, W., & Matthews, E. (2012). Changes in land surface water dynamics since the 1990s and relation to population pressure. *Geophysical Research Letters*, 39.

- Prigent, C., Rossow, W., & Matthews, E. (1997). Microwave land surface emissivities estimated from SSM/I observations. *Journal of Geophysical Research - Atmospheres*, 102, 21867–21890.
- Reichle, R., & Koster, R. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters*, 31.
- Ressources Naturelles Canada (2008). <http://www.rncan.gc.ca/sciences-terre/produits-services/produits-cartographie/geoscape/hydropanorama/bassin-riviere-bow/5930>, .
- Rudiger, C., Calvet, J., Gruhier, C., Holmes, T., de Jeu, R., & Wagner, W. (2009). An inter-comparaison of ERS-SCAT and AMSR-E soil moisture observations with model simulations over France. *American Meteorological Society*, 10, 431–447.
- Saleh, K., Kerr, Y., Richaume, P., Escorihuela, M., Panciera, R., Delwart, S., Boulet, G., Maisongrande, P., Walker, J., Wursteisen, P., & Wigneron, J. (2009). Soil moisture retrievals at L-band using a two-step inversion approach (COSMOS/NAFE'05 Experiment). *Remote Sensing of Environment*, 113, 1304–1312.
- Salvadori, G., & de Michele, C. (2007). On the use of copulas in hydrology : theory and practice. *Journal of Hydrologic Engineering*, 12, 369–380.
- Scherer, T., Seelig, B., & Franzen, D. (1996). *Soil, water and plant characteristic important to irrigation*. Technical Report National Dakota State University.
- Schlenz, F., dall'Amico, J., Loew, A., & Mauser, W. (2012). Uncertainty assessment of the SMOS validation in the upper Danube catchment. *IEEE Transactions on Geoscience and Remote Sensing*, 50, 1517–1529.
- Schmugge, T. (1985). *Remote sensing of soil moisture in hydrological forecasting*.
- Schweizer, B., & Wolff, E. (1981). On nonparametric measures of dependence for random variables. *Annals of statistics*, 9, 870–885.
- Scipal, K., Dorigo, W., & de Jeu, R. (2010). Triple collocation - a new tool to determine the error structure of global soil moisture products. *Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium*, (pp. 4426–4429).
- Serinaldi, F., & Grimaldi, S. (2007). Fully nested 3-copula : procedure and application on hydrological data. *Journal of Hydrologic Engineering*, 12, 420–430.
- Singh, V., & Strupczewski, W. (2007). Editorial. *Journal of Hydrologic Engineering*, 12, 345.
- Stoffelen, A. (1998). Toward the true near-surface wind speed : error modeling and calibration using triple collocation. *Journal of Geophysical Research*, 103, 7755–7766.
- Stéphanoise des Eaux (2007). http://www.stephanoise-eaux.fr/index.php?page=cadre1-2.php&article=cycle_eau&ss_menu=jeune&nom_menu=Tout%20sur%20l%27eau&ss_menu2=12, .
- Trivedi, P., & Zimmer, D. (2005). Copula modeling : an introduction for practitioners. *Foundations and Trends in Econometrics*, 1, 1–111.
- Ulaby, F., Moore, R., & Fung, A. (1986). *Microwave Remote Sensing - Active and Passive*.
- University of Kansas (1966). <http://mapas.owje.com/img/Mapa-de-la-Vegetacion-de-Estados-Unidos-2730.jpg>. *National Atlas - Potential Natural Vegetation*, .

- Van de Griend, A., & Owe, M. (1994). Microwave vegetation optical depth and inverse modeling of soil emissivity using Nimbus/SMMR satellite observations. *Meteorology and Atmospheric Physics*, 54, 225–239.
- Vincent, L., Zhang, X., Bonsal, B., & Hogg, W. (2002). Homogeneization of daily temperatures over Canada. *Journal of Climate*, 15, 1322–1334.
- Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sensing of Environment*, 70, 191–207.
- Wang, J., & Choudhury, B. (1981). Remote sensing of soil moisture content over bare field at 1.4 GHz frequency. *Journal of Geophysical Research*, 86, 5277–5282.
- Wang, J., & Schmugge, T. (1980). An empirical model for the complex dielectric permittivity of soil as a function of water content. *IEEE Transactions on Geoscience and Remote Sensing*, 18, 288–295.
- Wegmuller, U., Maetzler, C., & Njoku, E. (1995). Canopy opacity models. *Passive Microwave Remote Sensing of Land Atmosphere Interactions*, (pp. 375–387).
- Wigneron, J., Chanzy, A., Calvet, J., & Brugier, N. (1995). A simple algorithm to retrieve soil moisture and vegetation biomass using passive microwave measurements over crop fields. *Remote Sensing of Environment*, 51, 331–341.
- Wigneron, J., Chanzy, A., Kerr, Y., Lawrence, H., Shi, J., Escorihuela, M., Mironov, V., Mialon, A., Demontoux, F., de Rosnay, P., & Saleh-Contell, K. (2011). Evaluating an improved parameterization of the soil emission in L-MEB. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 1177–1189.
- Wigneron, J., Kerr, Y., Chanzy, A., & Jin, Y. (1993a). Inversion of surface parameters from passive microwave measurements over a soybean field. *Remote Sensing of Environment*, 46, 61–72.
- Wigneron, J., Kerr, Y., Chanzy, A., & Jin, Y. (1993b). Inversion of surface parameters from passive microwave measurements over a soybean field. *Remote Sensing of Environment*, 46, 61–72.
- Wigneron, J., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M., Richaume, P., Ferrazzoli, P., de Rosnay, P., Gurney, R., Calvet, J., Grant, J., Guglielmetti, M., Hornbuckle, B., Mätzler, C., Pellarin, T., & Schwank, M. (2007). L-band Microwave Emission of the Biosphere (L-MEB) model : description and calibration against experimental data sets over crop fields. *Remote Sensing of Environment*, 107, 639–655.
- Wigneron, J., Laguerre, L., & Kerr, Y. (2001). A simple parameterization of the L-band microwave emission from rough agricultural soil. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1697–1707.
- Wigneron, J., Parde, M., Waldteufel, P., Chanzy, A., Kerr, Y., Schmidl, S., & Skou, N. (2004). Characterizing the dependence of vegetation model parameters on crop structure, incidence angle, and polarization at L-band. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 416–425.
- Woodhouse, I. (2006). *Introduction to microwave remote sensing*. Taylor and Francis.

- World Meteorological Organization, Intergovernmental Oceanographic Commission, United Nations Environment Programme, & International Council for Science (2010). *Implementation plan for the global observing system for climate in support of the UNFCCC*. Technical Report GCOS-138 Global Climate Observing System.
- Zhang, L., & Singh, V. (2007). Trivariate flood frequency analysis using the Gumbel-Hougaard copula. *Journal of Hydrologic Engineering*, 12, 431–439.